A VARYING PARAMETER MODEL OF STOCK RETURNS

BY

YOUNG-HOON KOO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1992

To my parents, wife and children

ACKNOWLEDGEMENTS

I would like to express my heartfelt thanks to my supervisor, Dr. G. S. Maddala, for his warm encouragement and helpful suggestions which provided me with the tools and motivation necessary to complete this dissertation.

I would also like to thank Dr. David Denslow, Dr. Mark Rush and Dr. Robert Emerson for their valuable comments on this work. Also, many thanks to Dr. Adrian Pagan of the University of Rochester for providing me with the data sets essential to the empirical work of my dissertation.

This work could not have been produced without the patience and invaluable support which my wife, Yang-ok, has shown me over the years. I dedicate this work to her.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

A VARYING PARAMETER MODEL OF STOCK RETURNS

By

Young-Hoon Koo

August 1992

Chairperson: Dr. G. S. Maddala
Major Department: Economics

The economy consists of various dynamic activities whose interrelationships vary over time. There have been many attempts to account for the variations of stock returns with the movements of those dynamic activities. These activities include the growth rates of gross national products, changes in industrial production, inflation rates, short- and long-term interest rates, and yields on corporate bonds. Those time-varying relationships with stock returns can be represented by the coefficients in stock return generating equations. Most of the previous work, however, relies on the very restrictive assumption that the parameters in the equations remain constant over time. This restrictive assumption leads to the low explanatory power of the models on stock returns. In this work, I construct a more general model of stock returns using the Kalman filter which allows the parameters to change over time. In addition, I also

generalize an assumption on error terms by letting the variance-covariance matrices of the disturbance terms in both measurement and transition equations follow autoregressive conditional heteroskedasticity.

Before performing empirical work on the model, I check whether a general error distribution (GED), employed in Nelson's work, or a normal distribution fits the distribution of actual stock returns better and find that the performance of the GED distribution is almost indistinguishable from that of the normal distribution.

Under the assumption that the normal distribution fits the disturbance terms in the model, the null hypothesis that parameters remain constant is rejected at the 5 percent level of significance. I also find little evidence that stock markets are informationally inefficient.

Based on the results of estimation of the model, I find that the time-varying parameter model yields better explanatory power over alternative models of stock returns. I re-examine features of U.S. monthly stock returns.

CHAPTER 1
INTRODUCTION

So far there have been a great number of studies on how stock prices or stock returns are determined and how risks are priced. The capital asset pricing model (CAPM), introduced by Sharpe (1964), Litner (1965) and Mossin (1966), has been a major framework for analyzing financial markets for a long time. It states that expected returns from holding securities should be linearly related to the so-called market beta. But the CAPM uses assumptions that are too strong to be testable.

As a testable alternative to the CAPM, the arbitrage pricing theory (APT) has been proposed by Ross (1976) and developed by Roll (1977, 1981), Roll and Ross (1980), Connor (1980), Reinganum (1981), Shanken (1982), Chamberlain (1983), Chamberlain and Rothschild (1983), Chen (1984), Dhrymes, Friend and Gultekin (1984, 1985), Wei (1988), Mei (1989), Sawyer (1989), and others. The APT states that in equilibrium expected returns from holding securities should be linear combinations of factor loadings. In the APT models, however, explanatory factors are neither explicitly identified nor is the uniqueness of solutions, obtained using the factor analysis on the covariance matrix of stock returns, guaranteed.

Recently, many financial economists such as Black and Fischer (1976), Bollerslev (1986), French, Schwert and Stambaugh (1987), Nelson(1988a, b, c, 1989), Schwert (1989a, b, c), LeBaron (1990), Merville & Pieptea (1989), and Kearns and Pagan (1990) have attempted to relate the volatility of stock returns to the volatilities of major economic variables such as real GNP growth rate, the inflation rate, growth rates of industrial production, financial leverage, relative price variability and dividend yields. Many of them also implicitly or explicitly assume that stock returns are determined by the volatility of returns. In most of the previous work on volatility, stock volatility is measured by the variance of the error terms in models of stock returns. Note that the error term in a specific model is the part unexplained by the model. Thus if, besides the variance of stock returns such variables as described above are directly incorporated into the model of stock return, then the explanatory power of the model will be increased as long as they have some explanatory power over the market variance of stock returns.

In addition, most of the previous studies on stock returns mentioned above are static in the sense that they assume the constancy of the coefficients which relate the stock returns or the volatilities of the stock returns to explanatory economic variables. But this assumption appears to be less reasonable if we take into account the fact that the economy is a collection of dynamic activities whose relations vary

constantly.

Many of the studies on stock returns, like those by Engle and Bollerslev (1986), Kearns and Pagan (1990), Bollerslev, Engle & Wooldridge (1987), Bollerslev (1988), employ the linear GARCH model with normally distributed error terms. But Nelson (1989) shows that the distribution of stock returns tends to have fatter tails than the normal distribution. And to ensure nonnegativity of the variance these studies also impose very restrictive assumptions that all the coefficients in the return variance formulas are nonnegative. This rules out the possibility of oscillatory behavior of the return variance. Next, the GARCH models assume that the distribution of stock returns is symmetric so it cannot explain the negative skewness of stock returns or the tendency for market volatility to rise after the market declines (Campbell and Hentschel, 1990). Third, in conventional GARCH models, the current σ_t^2 is determined by a linear combination of lagged σ_{t-i}^2, $i = 1,2,...,p$, and squared terms of current and lagged errors. Thus it does not reflect whether unanticipated excess returns are positive or negative. But Schwert (1989a), Bollerslev (1987) and Nelson (1988a, 1989) show that stock returns are negatively correlated with changes in volatility of returns. Lastly, the GARCH models, in general, do not highlight the economic theory that explains the behavior of stock returns in equilibrium.

In this paper, using the Kalman filter, I build a more

dynamic model in which the coefficients of the return generating structure and their conditional covariances are allowed to vary over time. In addition, the variances of disturbances in the state space model, conditional on the past observations, are assumed to follow ARCH(1) processes. This type of model can account for phenomena in the dynamic economy in a more proper way than the static models can. Especially, in this work, the covariance matrix of disturbances in a transition equation, conditional on past covariance matrix, is determined by a linear combination of a matrix of constants and a cross product of past disturbance term given by the Eq (4.1.5). It will be called a matrix ARCH (MARCH). There has been little work on a MARCH model in econometrics so this paper can be regarded as a frontier in this area. And most of the previous work, such as Harvey and Ruiz (1990), which applies a Kalman filter to stock return or price data assumes the case where there is only one explanatory variable and the transition coefficient is already known to be equal to one. But these are very restrictive assumptions in the sense that, as can be seen in Chapter 2, there are several other variables besides one-lagged stock return, for example, real GNP, industrial production, investment, inflation rate, yield on bonds, short-term interest rate, and so on, which have some explanatory power over stock returns; moreover, the coefficients on those variables are seldom equal to one. So in this work I am going to construct a more general model in

which some of those macroeconomic variables described above are used as explanatory variables and the transition matrix of the parameters is not restricted to being equal to an identity matrix.

The structure of this work is as follows. In Chapter 2, I summarize the characteristics of the U.S. stock returns: volatility of stock returns, relationship of volatility to economic activity, and behavior of stock returns, which have been found in the previous work. In Chapter 3, the performance when stock returns are assumed to have a general error distribution will be compared with that when stock returns are assumed to have a Gaussian distribution. The comparison between the two distributions will be performed by a naive estimator and a kernel estimator of density, which will be estimated with new methods. In Chapter 4, I construct a more dynamic model in terms of a combination of a measurement equation and a transition equation in which disturbance terms conditional on the past disturbances are normally distributed with means zero and variances following an ARCH(1) process. Then I am going to derive the Kalman filter of the model in which the variances of the disturbance terms follow an ARCH process and then present the procedures of mathematical optimization of the unknown parameters in the model when using a maximum likelihood estimation method. In Chapter 5, I will estimate a simpler version of model in which the variances of error terms do not follow an ARCH process

since estimation of the more general version involves very high computational cost. Also, I will compare the performance of the model in this work with those of alternative models of stock returns and then perform tests on parameter constancy and stock market efficiency. In Chapter 6, based on the results of estimation in the previous chapter, I will attempt to reinterpret some of the important features of stock returns found in Chapter 2. Chapter 7 will presents some conclusions.

CHAPTER 2
CHARACTERISTICS OF THE U.S. STOCK RETURNS

In this chapter, I am going to summarize the results of the previous work on the behavior of stock returns. Some of them will be reinterpreted in terms of the outcomes obtained from our model.

2.1 Volatility of Stock Returns

First, the volatility of returns of stocks and bonds varies over time (French, Schwert, and Stambaugh, 1987; Schwert, 1989a; Merville and Pieptea, 1989; Turner, Startz and Nelson, 1989). In particular, the volatility tends to be generally very high during war, economic recession, oil shocks, and banking or financial panic (Schwert, 1989a). But strong elastic forces act on the stock volatility around its long-term value (Merville and Pieptea, 1989).

Second, shocks to the volatility of stock returns are persistent (Nelson, 1989). It follows that volatilities are highly autocorrelated so lagged volatility of returns has more predictive power on contemporaneous volatility than other variables such as lagged volatility of inflation (PPI), bond returns, interest rates, and monetary base (Schwert, 1989a).

In other words, autocorrelation decays very slowly even though autocorrelation $\rho(\sigma_{t+s}, \sigma_t)$ is a decreasing function of the time lags (Merville and Pieptea, 1989). It follows that shocks have both permanent and transitory components (Schwert, 1989b).

Third, in the short run, variation from the mean is primarily determined by transitory movements, while in the long run variation from the mean is primarily determined by both permanent and transitory movements (Merville and Pieptea, 1989).

Fourth, a non-trading day contributes much less to volatility than a trading day (French and Roll, 1986; Nelson, 1989d). Trading activity does not explain much of the variation in volatility through time but trading volume does (Schwert, 1989a).

Fifth, the volatility of stock returns is higher than that of bond returns and interest rates and the estimates of volatility from daily data have much less error than those from monthly data (Schwert, 1989a).

2.2 Relationship of Volatility to Economic Activity

First, expected returns vary over time (French, Schwert & Stambaugh, 1987; Campbell & Shiller, 1988; Fama & French, 1988a, b; Turner, Startz & Nelson, 1989; Ball & Kothari, 1989). The negative autocorrelation of stock returns can be explained largely by the time-varying expected returns, which

in turn are caused by variation in expected returns on the market portfolio, in relative risks of firms's investments, and in leverage (Ball & Kothari, 1989). On the other hand, French, Schwert, and Stambaugh (1987) showed that excess returns are positively related to the predictable volatility of stock returns while they are negatively related to the unexpected change in the volatility of stock returns.

Second, a large fraction of annual stock return variation can be accounted for by future real activities such as real GNP, industrial production, and investment (Fama, 1990). The future growth rate of industrial production has a especially significant and positive effect on the current real stock returns (Kaul & Seyhun, 1990; Shanken & Weinstein, 1990). Fama and French (1988b), however, found that earnings and dividend policy have relatively low predictive power on the stock return volatility. In addition, Fama (1990) showed that real activities explain more return variation for longer return horizons.

Third, declines in the stock market tend to be associated with subsequent increases in volatility (Nelson, 1989; Schwert, 1989a; Turner, Startz & Nelson, 1989).

2.3 Behavior of Stock Returns

First, the distribution of log stock returns tends to have fatter tail and more pronounced peak than the normal (Nelson, 1989; Bollerslev, 1987; Turner, Startz & Nelson, 1989).

Second, the distribution of log stock returns is negatively skewed, i.e., large negative log returns are more common than large positive returns; see Kearns and Pagan (1990), Campbell and Hentschel (1990). And the negative skewness increases with the conditional variance of the stock returns (Campbell & Hentschel, 1990).

Third, during the post-war period, stock returns are negatively related to inflation, to be more precise, relative price variability. It follows that stock returns are positively related to future real activity such as the growth rate of industrial production. The negative effects of relative price variability on the future output and thus on the contemporaneous real stock returns are largely caused by the supply shocks in the 1970s (Kaul & Seyhun, 1990).

Fourth, stock returns are predictable (Fama & French, 1988a; French, Schwert & Stambaugh, 1987). In particular, returns are more predictable for portfolios of small firms than those of large firms since in the small firm portfolio stationary components of stock prices tend to prevail over random components (Fama & French, 1988a). On the other hand, Poterba and Summers (1988) argue that stock returns are positively autocorrelated over short horizons and negatively autocorrelated over long horizons. Not surprisingly, long-horizon returns are more predictable since the slowly mean-reverting component of stock prices tends to induce negative correlation in return and slow mean reversion can be more

evident in the long-horizon returns (Fama & French, 1988a; Ball & Kothari, 1989). Fama and French (1988a) ascribe the negative autocorrelation of long-horizon returns to a common macroeconomic phenomenon to firm-specific factors.

Fifth, Fama (1990) found that dividend yield, default spread, and term spread are positively related to expected returns, respectively. But the unusually large realization of dividend news, either good or bad, will have a negative effect on the stock returns (Campbell & Hentschel, 1990).

CHAPTER 3
GED OR NORMAL DISTRIBUTION?

In his paper (1989), Nelson assumes that stock returns follow General Error Distribution (GED). In other words, the probability density function of stock returns(y) is given by

$$f(y) = \left[\Gamma\left(1 + \frac{1+\beta}{2}\right) 2^{1+\frac{1+\beta}{2}} \phi\right]^{-1} \exp\left[-\frac{1}{2}\left|\frac{y-\mu}{\phi}\right|^{\frac{2}{1+\beta}}\right]$$

where μ is the mean of stock returns, β is a tail-thickness parameter with its range $-1 < \beta \leq 1$, and ϕ is a positive number. If $\beta = 0$, then the stock returns are normally distributed. If $0 < \beta \leq 1$, then the distribution of y has the thicker tail than a normal distribution. If $-1 < \beta < 0$, the distribution of y has the thinner tail than a normal distribution.

To determine which distribution fits the empirical distribution better, we need to estimate the density function of distribution of stock returns.

3.1 Methods for Estimating the Empirical Density

Up to now, there have been many methods proposed for estimating the empirical density function. Histograms, so-called naive estimators (Fix & Hodges, 1951; Rosenblatt,

1956), kernel estimators, the nearest neighbor method (Loftsgaarden & Quesenberry, 1965), the variable kernel method (Breiman, Meisel & Purcell, 1977), orthogonal series estimators (Cencov, 1962), maximum penalized likelihood estimators (Good & Gaskins, 1971), general weight function estimators (Whittle, 1958), the reflection and replication techniques (Boneva, Kendall & Stefanov, 1971), and the transformation techniques (Copas & Fryer, 1980) are well known methods for density estimation.

3.1.1. Histograms

One of the simplest ways of estimating the density function is to draw the histograms of the sample. To construct the histogram, we have to split the entire range, which sufficiently covers the range of the sample data, into some suitable number of bins x by a certain size of bin width h. Then observing the number of observations in each bin and using the following formulation, we can obtain the density estimator :

$$\hat{f}(x) = \frac{1}{Nh}\ (no.\ of\ X_i \in bin\ x)$$

where $\{X_1, X_2, \cdots, X_N\}$ is a sample of N real observations and $f(\cdot)$ is the density function to be estimated. Note that it is the choice of bin width h that primarily determines the smoothness of the histogram.

3.1.2. The Naive Estimator and the Least Square Estimator Using Naive Estimation Data

The naive estimator of density function is a little more improved way of estimating density than the histogram method, using the definition of a probability density function f() :

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x-h < X < x+h) \qquad (3.1.1)$$

Then the naive estimator of density is given by

$$\hat{f}(x) = \frac{1}{2Nh} [no.\ of\ X_i \in (x-h, x+h)] \qquad (3.1.2)$$

The methods, however, such as the histograms and the naive estimator have some drawbacks as below. First, it is not a continuous function so that it has jumps at the points $X_i \pm h$ and has zero derivatives everywhere else. Second, the choice of bin width is made somewhat arbitrarily so that the smoothness of the density function can be somewhat misleading.

In order to overcome the problems, instead of using directly the naive estimator given by the Eq (3.1.2) as the estimator of true density, I am going to use it as a kind of transformed data which will be used for fitting the true density with a Gaussian distribution or a GED distribution. Then the problems described above almost disappear. Let us investigate which distribution out of the GED and the normal distribution fits the empirical distribution better when the non-linear least squares method (NLS) is applied. The main idea of this new method is to find optimal estimators of

parameters in the density functions which minimizes the mean squared error (MSE) of the distribution from the empirical distribution. For this purpose, the whole range between the lowest and the highest value of stock return will be split into some intervals,[1] and then the estimator of the density for each interval will be obtained by applying Eq (3.1.2). The difference in the values of the empirical distribution $f(y_i)$ and the theoretical distribution (GED or Normal) corresponding to the center of each interval will be computed and, then, the sum of squared differences will be minimized with respect to the unknown parameters δ in the theoretical distributions $h(y_i)$.

$$\underset{\delta}{Min} \quad \sum_{i=1}^{N} [\hat{f}(y_i) - h(y_i)]^2 \qquad (3.1.3)$$

where N is the number of intervals. In addition, the density function of the normal distribution to be estimated is given by

$$P(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] \qquad (3.1.4)$$

The density estimator obtained from this method will be called the least square estimator using the naive estimation data.

---

[1] The U.S. stock return data shows that the lowest and highest values of stock returns are -0.28794 and 0.37681, respectively. The distance of each interval will be 0.001 if the whole length between -0.295 and 0.395 is divided into 690 intervals.

3.1.3. The Kernel Estimator and the Pseudo-data Maximum Likelihood Method with a Kernel

Let's continue comparing the performance of two assumptions imposed on the distribution of stock returns using the kernel estimator of density which is one of the most commonly used estimators of density. The kernel estimator with kernel K is given by

$$\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right) \tag{3.1.5}$$

where

$$\int_{-\infty}^{\infty} K(t)\, dt = 1 \tag{3.1.6}$$

and h is the smoothing parameter. The kernel estimator can be interpreted as a sum of 'bumps' placed at the observations. Two major properties of kernel estimates should be noted here. One is that the estimator given by the Eq (3.1.5) will be a probability density if the kernel K() is everywhere non-negative and satisfies the Eq (3.1.6). The other is that it will be continuous and differentiable if K() is continuous and differentiable. In addition, when applied to data from long-tailed distributions, there is a tendency for spurious noise to appear in the tails of the estimates since the window width is fixed across the entire sample. Of course, the kernel estimator depends on what kind of kernel function is used, what size of window width h is used, and, in turn, how the optimal size of h is chosen. The first problem is to

determine what kind of kernel function should be used. The Gaussian and GED kernels are summarized in Table (3.1.1). In general, the choice of kernel function is made subjectively according to author's preference.

The problem is how to determine the optimal magnitude of the window width h. A solution to the problem is to choose the optimal h which minimizes the mean integrated squared error of estimated density from the true density:

Table 3.1.1 Kernel Functions

Kernel	Kernel Function
Gaussian	$K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} t^2\right)$
GED	$K(t) = \dfrac{\delta \exp\left[-\frac{1}{2}\left\lvert\frac{t}{\lambda}\right\rvert^{\delta}\right]}{\lambda 2^{1+\frac{1}{\delta}} \Gamma\left(\frac{1}{\delta}\right)}$
	$\lambda = \left[\dfrac{\Gamma(1/\delta)}{2^{\frac{2}{\delta}}\Gamma(3/\delta)}\right]^{\frac{1}{2}}$

$$\underset{\delta,h}{Min} \int E[(f-\hat{f})^2]\,dx \qquad (3.1.7)$$

where δ is unknown parameter in the density function. The mean sqaured error can be rewritten as sum of sqaure of bias and variance. First, the bias is given by

$$bias_h(x) = f(x) - E[\hat{f}(x)] = f(x) - \int \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y)\,dy \qquad (3.1.8)$$

In the above equation, it is not too hard to find out that the bias in the estimation of f(x) depends on the window width h, not directly on the sample size, as well as on the kernel K. Suppose that the kernel k() is a symmetric function satisfying

$$\int K(t)\,dt = 1, \quad \int tK(t)\,dt = 0, \quad \int t^2K(t)\,dt = k_2\,(\neq 0) \qquad (3.1.9)$$

and that the unknown density f() is continuously differentiable of all orders. Then letting y = x - th and using a Taylor series expansion on f(x - th), the bias can be rewritten as

$$\begin{aligned} bias_h(x) &= hf'(x)\int tK(t)\,dt - \frac{1}{2}h^2f''(x)\int t^2K(t)\,dt + \cdots \\ &= \frac{1}{2}h^2f''(x)\,k_2 + higher\text{-}order\ terms\ of\ h \end{aligned} \qquad (3.1.10)$$

In a similar way, the variance of the density estimate can be approximated by

$$\begin{aligned} var\,\hat{f}(x) &= \frac{1}{n}\int \frac{1}{h^2}K\left(\frac{x-y}{h}\right)^2 f(y)\,dy - \frac{1}{n}[f(x) + bias_h(x)]^2 \\ &\approx \frac{1}{nh}\int (x-ht)K(t)^2dt - \frac{1}{n}[f(x) + O(h^2)]^2 \\ &\approx \frac{1}{nh}\int [f(x) - htf'(x) + \cdots]K(t)^2dt + O(n^{-1}) \\ &\approx \frac{1}{nh}f(x)\int K(t)^2dt \end{aligned} \qquad (3.1.11)$$

It follows from the above two equations that the approximate mean integrated squared error is given by

$$\frac{1}{4}h^4k_2^2\int f''(x)^2dx + \frac{1}{nh}\int K(t)^2dt \qquad (3.1.12)$$

Then the optimal magnitude of h can be obtained by minimizing the approximate mean integrated squared error of the density estimate with respect to h as follows :

$$h_{opt} = k_2^{-2/5}\left[\int K(t)^2 dt\right]^{1/5}\left[\int f''(x)^2 dx\right]^{-1/5} n^{-1/5} \qquad (3.1.13)$$

As can be seen in the Eq (3.1.13), this method produces some useful properties of the optimal h. First, h_{opt} converges to zero as n increases, but at a very slow rate. Second, smaller values of h will be appropriate for more rapidly fluctuating densities since such densities have larger absolute values of $f''(x)$.

A critical drawback of this method, however, is that the true density function is unknown at the stage where the density function is being estimated. Also frequent use of approximation of the objective function will lead to loss of information, in other words, inaccuracy of mathematical calculation of the ideal window width. And the assumption of symmetricity of kernel function is also somewhat restrictive. So an alternative method to overcome the problem should not rely on the true density. At this moment, maximum likelihood estimation technique will be very useful. The estimation procedures are as follows. First, generate an arbitrary series, $\{x^j,\ j = 1,2,\cdots,M\}$, whose range is sufficiently larger than that of the sample data. Second, as in the conventional maximum likelihood estimation, construct the log-likelihood function using the Eq (3.1.5) over the pseudo-data

which sufficiently covers the range of the sample data.

$$L = \ln \prod_{j=1}^{M} \hat{f}(x^j)$$
$$= \sum_{j=1}^{M} \left[-\ln(nh) + \ln \sum_{i=1}^{n} K\left(\frac{x^j - x_i}{h}\right) \right] \qquad (3.1.14)$$

Third, taking the partial derivatives of the log-likelihood function with respect to h and unknown parameters δ, we will obtain the optimal value of the window width h which satisfies

$$\frac{\partial L}{\partial h} = -\frac{M}{h} + \sum_{j=1}^{M} \frac{\sum_{i=1}^{n} \frac{\partial K\left(\frac{x^j - x_i}{h}\right)}{\partial h}}{\sum_{i=1}^{n} K\left(\frac{x^j - x_i}{h}\right)} \qquad (3.1.15)$$
$$= 0$$

Suppose that K() is a Gaussian kernel function, that is,

$$K\left(\frac{x^j - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x^j - x_i}{h}\right)^2\right] \qquad (3.1.16)$$

Then denoting $K((x^j - x_i)/h)$ as K_{ij} and taking the partial derivative of the kernel function with respect to h,

$$\frac{\partial K_{ij}}{\partial h} = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x^j - x_i}{h}\right)^2\right] \frac{(x^j - x_i)^2}{h^3}$$
$$= K_{ij} \frac{(x^j - x_i)^2}{h^3} \qquad (3.1.17)$$

Substituting the Eq (3.1.17) back into the Eq (3.1.15), then we will obtain the optimal window width, h_{opt}, as below :

$$h_{opt} = \left[\frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{n} w_{ij}(x^j - x_i)^2\right]^{1/2} \tag{3.1.18}$$

where

$$w_{ij} = \frac{K_{ij}}{\sum_{i=1}^{n} K_{ij}}, \qquad \sum_{i=1}^{n} w_{ij} = 1,$$

This will be called the pseudo-data maximum likelihood method with the kernel since the kernel estimator of density is estimated by applying the maximum likelihood technique to the psuedo-data. The optimal window width obtained from using the maximum likelihood estimation method has some useful properties. First, unlike the minimization method of the mean integrated squared error, the optimal h does not require any information concerning the true density function, but relies on the kernel function. This is a very satisfactory result since it sounds contradictory to assume that the true density is known while the true density is under estimation. Assume that $|x^j - x_i| < 1$. Then as the sample size n increases, given x^j, the denominator of the Eq (3.1.18) expands faster than its numerator does, since $(x^j - x_i)^2 < 1$. It follows that the optimal window width h will converge to zero as the sample size increases. This is compatible with the result obtained from the previous method. Third, the size of the arbitrary series generated for construction of likelihood function can affect the optimal h in different directions. Suppose that

$$\sum_{j=1}^{M}\sum_{i=1}^{n} w_{ij}(x^j - x_i)^2 \sim O(M^s)$$

If s is greater (or less) than 1, the ideal window width will increase (or decrease) as the size of an arbitrary series M increases. If s equals to 1, the optimal h will not be affected by changes in M. Lastly, the optimal h can be interpreted as weighted average of the distances between observations of the sample and points generated for construction of likelihood function.

Next, suppose that K() is a GED kernel function, that is,

$$K\left(\frac{x^j - x_i}{h}\right) = \frac{\delta \exp\left[-\frac{1}{2}\left|\frac{x^j - x_i}{h\lambda}\right|^{\delta}\right]}{\lambda 2^{1+\frac{1}{\delta}}\Gamma\left(\frac{1}{\delta}\right)} \tag{3.1.19}$$

where

$$\lambda = \left[\frac{\Gamma(1/\delta)}{2^{\frac{2}{\delta}}\Gamma(3/\delta)}\right]^{\frac{1}{2}}$$

Taking the partial derivative of the Eq (3.1.19) with respect to h, we will obtain

$$\frac{\partial K_{ij}}{\partial h} = K_{ij}\left[\frac{1}{2}\left|\frac{x^j - x_i}{\lambda}\right|^{\delta} h^{-(\delta+1)}\right] \tag{3.1.20}$$

Substituting the Eq (3.1.20) into the Eq(3.1.15) and solving it over h, we will get the following solution :

$$h_{opt} = \left[\frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{n} P_{ij}\left(\frac{1}{2}\left|\frac{x^j - x_i}{\lambda}\right|^{\delta}\right)\right]^{\frac{1}{\delta}} \tag{3.1.21}$$

where

$$P_{ij} = \frac{K_{ij}}{\sum_{i=1}^{n} K_{ij}}$$

Some conclusions similar to the case with a Normal kernel can be drawn. First, the optimal window width h does not require any information concerning the true density function. Second, as the sample size increases, given x^j, the optimal h will converge to zero since $|x^j - x_i|$ is assumed to be less than one. Note the speed of convergence to zero will depend upon the value of tail-thickness parameter δ. Third, the optimal window width obtained under a Gaussian kernel can be regarded as a special case of that obtained under a GED kernel, in which $\delta = 2$: see the Eq (3.1.18) and the Eq (3.1.21).

3.1.4. The Nearest Neighbor Method

The idea in the nearest neighbor method is to adapt the degree of smoothing to the local density of data. The degree of smoothing is controlled by an integer k which is considerably smaller than the sample size. Let $d(t,x_i)$ denote the distance between point t and the i-th point of the sample, x_i and assume that

$$d(t,x_1) \leq d(t,x_2) \leq \cdots \leq d(t,x_n)$$

Then the k-th nearest neighbour density estimate is given by

$$\hat{f}(t) = \frac{1}{nd(t,x_k)} \frac{k-1}{2} \qquad (3.1.22)$$

The k-th nearest neighbour estimate has two critical drawbacks : one is that it is not a smooth curve while the other is that it will not itself be a probability density since it does not integrate to one. Thus the nearest neighbour estimate may not be appropriate if an estimate of the entire density is required.

It is possible to generalize the k-th nearest neighbour density estimate by applying the kernel estimator. The generalized k-th nearest neighbour density estimate is defined by

$$\hat{f}(t) = \frac{1}{nd(t,x_k)} \sum_{i=1}^{n} K\left(\frac{t-x_i}{d(t,x_k)}\right) \tag{3.1.23}$$

3.1.5. The Variable Kernel Method

The variable kernel estimate is obtained through an application of the kernel method to the generalized κthe nearest neighbour method and the kernel method. The variable kernel estimate of density with smoothing parameter h is defined by

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{hd_{i,k}} K\left(\frac{t-x_i}{hd_{i,k}}\right) \tag{3.1.24}$$

where K() is a kernel function, κ is a positive number, $d_{i,k}$ is the distance from x_i to the κth nearest point in the set including the other n-1 data points.

Some important features of the variable kernel method are as

follows. First, the variable kernel estimate is a probability density function if the kernel is. Second, like the kernel estimate, it will preserve all the local smoothness properties of the kernel. Third, the window width of the kernel placed on the point x_i is proportional to $d_{i,k}$, so that the data point in the region where the data are sparse will have flatter kernels. Fourth, for a given κ, the degree of smoothing depends on h, and the reponsiveness of the window width choice to each data point depends on κ.

3.1.6. Orthogonal Series Estimators

The orthogonal series estimation method is a way of estimating density function f() by estimating the coefficients of Fourier expansion of the data sequences. Suppose that we estimate a density function f() of the sequence $\phi_v(x)$, which is a function of random variable x, over some range, say [a, b]. And assume that $\{\phi_v\}$ is orthogonal on [a, b]. Then the vth Fourier coefficient of f relative to $\{\phi_v\}$ is given by

$$f_v = \int_a^b f(x)\,\phi_v(x)\,dx \qquad (v = 1, 2, 3, \cdots) \qquad (3.1.25)$$

The Eq (3.1.25) can be rewritten as

$$f_v = E[\phi_v(x)]$$

so that an unbiased estimator of f_v based on {x} from f is given by

$$\hat{f}_v = \frac{1}{n}\sum_{i=1}^{n} \phi_v(x_i)$$

3.1.7. Maximum Penalized Likelihood Estimators

It is possible to get better estimators of a density by adjusting the conflict between smoothness and goodness-of-fit to the data. A maximum penalized likelihood estimate is obtained by incorporating into the likelihood function a term which represents the roughness of a density. The penalized likelihood of a curve g based on observations, $\{X_i,\ i = 1, 2, \cdots, n\}$, is given by

$$L_\alpha(g) = \sum_{i=1}^{N} \log g(X_i) - \alpha R(g)$$

where α is a positive smoothing parameter and R(g) measures the roughness of g. Then the maximum penalized likelihood estimator of a density function g be evaluated at estimated parameters which maximize $L_\alpha(g)$ under the constraints that

$$\int_{-\infty}^{\infty} g(x)\,dx = 1 \ ,\ g(x) \geq 0 \text{ for all } x, \text{ and } R(g) < \infty.$$

As a result, it follows that the maximum penalized likelihood estimator will be a probability density.

3.1.8. General Weight Function Estimator

The density estimators discussed above can be expressed in a more general form. That is, the general weight function estimator of a density will be given by

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} W(X_i, x)$$

where

$\int_{-\infty}^{\infty} W(X_i, x)\, dx = 1$ and $W(X_i, x) \geq 0$ for all X_i and x. The examples are summarized as below :

Table 3.1.2 Several Weight Functions

Method	Weight function
Histogram	$1/h(x)$ if X_i and x fall in the same bin 0 otherwise
Kernel Estimator	$(1/h)K((x - X_i)/h)$
Orthogonal Series Estimator	$\sum_{v=0}^{K} \phi_v(X_i)\phi_v(x)$

3.2 Comparison between Performances of GED and Normal

In this section, the performance of the GED distribution will be compared with that of the normal distribution by two methods discussed above, more specifically, the least sqaure estimator using the naive estimation data and the pseudo-data maximum likelihood estimation using the kernel. The reasons for choosing these methods are that it is relatively easy to estimate them and that they shows some desirable properties as density estimators.

3.2.1. Empirical Results of the Least Square Method Using the Naive Estimation Data

The theoretical distributions are estimated for 3 different sizes of intervals. The results of estimation of density functions are summarized in Table (3.2.1) and Table (3.2.2). Table (3.2.1) shows that the mean of stock returns is approximately 1.1% while the tail-thickness variable ß is, though close to 0, approximately 0.14. It follows that the distribution of stock returns will have slightly thicker tails than the normal distribution. But it does not make a big difference between the GED and the Normal since the estimates of ß are close to 0 in all cases. As can be seen in Table (3.2.3) and Figure (3.2.1), the difference between the two distributions is almost indistinguishable except that the GED has slightly higher kurtosis and thicker tails than the normal. The relative advantage of the GED is less than 1% in

terms of the magnitude of the estimates of mean squared errors whereas it involves a big computational cost. Especially, the absolute value contained in the GED density function tends to interrupt optimization procedures since the sign of the gradient vector becomes very unstable due to it. Taking into account both the relative advantage and the computational costs when the GED distribution rather than a Gaussian distribution is used, it is very hard to say that the GED distribution is much more appropriate as the distribution of the monthly stock return indexes of the U.S.A. than a Gaussian distribution. Another reason why I decide to employ the GED as the stock return distribution is that unlike a normal distribution a linear combination of the series which are GED-distributed is not, in general, GED-distributed. So it will be very much complicated to derive a Kalman filter using the GED distribution.

Table 3.2.1 Results of Estimation of GED

Interval	ϕ	β	μ
0.0001	0.03671 (0.00030)	0.13903 (0.01245)	0.01075 (0.00015)
0.0005	0.03673 (0.00066)	0.13856 (0.02785)	0.01114 (0.00034)
0.001	0.03678 (0.00093)	0.13651 (0.03931)	0.01163 (0.00048)

( ) : standard error of coefficient estimate

Table 3.2.2 Results of Estimation of Normal Distribution

Interval	σ	μ
0.0001	0.03975 (0.00013)	0.01093 (0.00015)
0.0005	0.03976 (0.00028)	0.01132 (0.00033)
0.001	0.03976 (0.00040)	0.01181 (0.00047)

( ) : standard error of coefficient estimate

Table 3.2.3 Comparison between Estimated Mean Squared Errors

Interval	MSE1(NORMAL)	MSE2(GED)	MSE2/MSE1
0.0001	8.88284	8.87300	0.99889
0.0005	1.99013	1.98035	0.99509
0.001	1.05450	1.04599	0.99099

3.2.2. Empirical Results of Estimation of a Density Function Using the Pseudo-data MLE Method with the Kernel

The optimal window width when we use a Gaussian Kernel estimating the density function of stock returns by a MLE method are summarized in Table (3.2.4). When the range covering sufficiently the maximum and the minimum of stock returns is divided into 100 points, the optimal window width is 0.28682 and is significantly different from zero at 5 percents of sinificance level. The same optimization is applied to different number of pseudo-sample points. Table (3.2.4) shows that there appears to exist a globally optimal window width around 120 of pseudo sample points since the value of an average likelihood attains the maximum at M equal

to 120.

Table 3.2.4 Optimal Window Width with Normal Kernel

M	Window Width	Standard Error	Average Likelihood
100	0.28682	0.02088	-0.184729
120	0.28342	0.01885	-0.173143
150	0.28352	0.01686	-0.173484
200	0.28362	0.01461	-0.173830
300	0.28372	0.01193	-0.174181
500	0.28381	0.00924	-0.174466

On the other hand, the optimal tail-thickness variables δ and window width h_{opt} with a GED kernel function are estimated using a Gauss program. Originally, the estimate of δ was expected to be within the range of 2.0 ± 1.0. When the estimation is performed over 100 different points as in a Gaussian kernel case, however, the estimate of δ turns out to be equal to 228.67104 with a standard error of 2094.89694. This implies that it is statistically insignificant. Probably, such a result arises from the fact that unlike a Gaussian distribution case, a linear combination of the GED distributions does not necessarily lead to a new GED distribution. But it gives the estimate of optimal window width which is very close to that obtained under the Gaussian kernel. That is, the estimate of h_{opt} is equal to 0.28691 with a standard error of 0.00836 so it is statistically significant. It implies that there is no big difference between the estimates of the optimal window width obtained

under both the Gaussian and GED kernels. The results of estimation using the GED kernel function are summarized in Table (3.2.5).

Of course, we attempted estimation in these cases where M is greater than 100, but the programs for the estimation of the optimal window width and tail-thickness parameters never converged.

Table 3.2.5 Optimal Window Width with GED (M = 100)

Parameter	Estimate	S.E.	Ratio	Prob-value
Tail-thickness	228.67104	2094.89694	0.10916	0.91308
Window Width	0.28691	0.00836	34.30180	0.00000

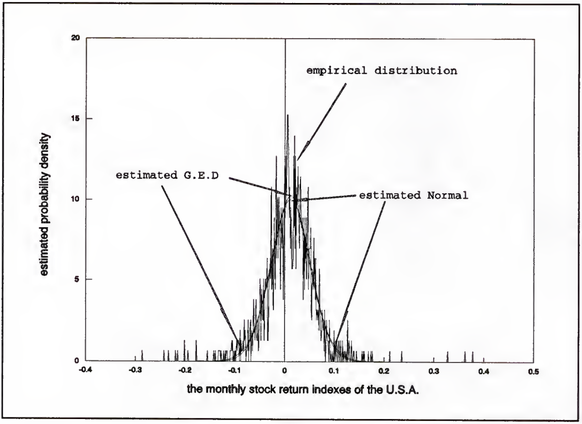

Figure 3.2.1 Comparison of GED with Normal Distribution

CHAPTER 4
DYNAMIC SPECIFICATION OF A MODEL USING THE KALMAN FILTER WITH ARCH DISTURBANCES

According to the theory in finance, the stock price in equilibrium is equal to the sum of discounted future dividends which will be paid to its holders. That is,

$$P_t = \sum_{t=1}^{\infty} \frac{DIVIDEND_t}{(1 + \delta_1)\cdots(1 + \delta_t)}$$

where δ_t is a discount factor at time t. Dividend payments at time t will be made based on the net cash flows a firm obtains during a given time period and the profitability of its business in the future. The net cash flows which the firm obtains during the period will be largely determined by the prices and quantities of the goods it produces, labor costs and capital costs. Thus changes in the prices and quantities of its products and inputs and the profitability of its future business will lead to changes in the stock prices of the firm.

On the other hand, suppose that a specific firm with very good performance in its business issues corporate bonds with high yields. Then the bond issuing will play a role as a signal that the firm is planning very good projects and have a positive effect on the stock prices of the firm. The difference between yield on the Moody's Aa corporate bond and

the short-term interest rate reflects, to some extent, the business conditions of the whole economy. Thus ceteris paribus, stock returns are determined by the term spread, the inflation rates and the real activity such as the growth rates of GNP and industrial production. Let us assume that the stock return is a linear function of the term spread, the inflation rate and the growth rate of industrial production.

The economy is a dynamic system which adjusts itself incessantly in response to various impacts on it such as oil shocks, strikes, changes in the political system, changes in the tax regime, wars, changes in financial environment, innovations in technology and management, and so on. In its adjustment procedures, relative power between several economic forces changes over time. In a specific model, such relations between the economic forces can be accounted for by the magnitudes of coefficients of the model. Thus it seems to be more plausible to assume that the coefficients vary over time than to assume that they remain unchanged. For example, the oil shocks during the 1970s caused the prices of petroleum products, in turn, those of petrochemicals and firms' fuel costs to sharply rise. On the one hand, consumers tried to reduce their consumption of these products or to divert their expenditures to other products. On the other hand, firms also tried to divert economic resources to other sectors which were less dependent on petroleum, or to invent more efficient production systems. All of there factors can affect the

cashflows and of course the stock prices of the firms. But in general it will take a considerably long time for consumers and producers to fully adjust themselves to the new environment because of the difficulty in changing consumers' preferences, restricted budgets, loss associated with diversion of investment, or inefficient bureaucratic behaviors. Also such changes as described above occur continuously in many sectors and give different impacts to the whole economy. Thus it seems to be more sensible to assume that from the point of the whole economy the interrelationships between economic factors changes gradually over time rather than abruptly. a smoothly varying parameter model can be regarded as a general form of discretely varying parameter model.

In this chapter, by explicitly taking account of the time-varying relationships between explanatory variables, I will construct a more dynamic model of stock returns.

4.1 The State Space Model

The dynamic model described above can be expressed in the form of a state space model. The state space model is composed of two important equations, a measurement equation and a transition equation.

4.1.1. A Measurement Equation

$$y_t = X_t\beta_t + d_t + \epsilon_t, \quad t = 1,2,\cdots,T \tag{4.1.1a}$$

where y_t is a Nx1 multivariate time series vector of stock returns for holding stocks from time t-1 to t, X_t is an Nxm matrix of explanatory variables, β_t is a mx1 vector, known as the state vector, and d_t is an Nx1 vector. And assume that ϵ_t is an Nx1 vector of serially uncorrelated disturbances with conditional mean zero and conditional covariance matrix H_t, that is,

$$E[\epsilon_t] = 0 \quad \text{and} \quad Var[\epsilon_t] = H_t \tag{4.1.1b}$$

where $H_t = H(\epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_{t-p})$

As the first attempt, consider a simple case where y_t is a univariate series with error terms following an ARCH(1) process and d_t is ignored. The variance of error conditional on the past errors, is determined by a linear combination of constant and squared error observed at time t-1.

$$y_t = x_t\beta_t + \epsilon_t \tag{4.1.2a}$$

$$\epsilon_t = h_t^{1/2} z_t \tag{4.1.2b}$$

$$\{z_t, -\infty < t < \infty\} \sim \text{iid } N[0, 1] \tag{4.1.2c}$$

$$h_t = q + r*\epsilon_{t-1}^2 \tag{4.1.2d}$$

where x_t is the 1xm vector of explanatory variables while β_t is the mx1 vector of coefficients. If ϵ_{t-1} is fixed and known at time t-1, the distribution of ϵ_t conditional on past observations will be normal with mean zero and variance h_t. However, such a case rarely happens in empirical studies.

Thus it should be noted that the distribution of ϵ_t conditional on past observations is symmetric but it is not, in general, normal since knowledge of past observations does not necessarily imply knowledge of past disturbances. For simplicity, we will assume that the distribution of ϵ_t conditional on past observations is normal with mean zero and variance $E_{t-1}[\epsilon^2_t]$. Let Ω_{t-1} denote a set of information on observations up to and including y_{t-1}, i.e, $\Omega_{t-1} = \{y_{t-1}, y_{t-2},\cdots,y_1\}$. And let b_{t-1} and P_{t-1} denote the optimal estimates of β_{t-1} and the mxm covariance matrix of b_{t-1} conditional on Ω_{t-1}, respectively. Taking the expectation conditional on Ω_{t-1} on both sides of the Eq (4.1.2d),

$$E_{t-1}[\epsilon^2_t] = q + r * E_{t-1}[\epsilon^2_{t-1}]. \tag{4.1.3}$$

$\epsilon_{t-1} = y_{t-1} - x_{t-1}\beta_{t-1}$ from the Eq (4.1.2a)

$$= (y_{t-1} - x_{t-1}b_{t-1}) + x_{t-1}(b_{t-1} - \beta_{t-1}) \tag{4.1.4a}$$

$$E_{t-1}[\epsilon^2_{t-1}] = e^2_{t-1} + x_{t-1}P_{t-1}x_{t-1}' \tag{4.1.4b}$$

where $e_{t-1} = y_{t-1} - x_{t-1}b_{t-1}$ and $P_{t-1} = E_{t-1}[(b_{t-1} - \beta_{t-1})(b_{t-1} - \beta_{t-1})']$. Substituting the Eq (4.1.4b) into the Eq (4.1.3) gives the variance of ϵ_t conditional on the information set available at time t-1,

$$E_{t-1}[\epsilon^2_t] = q + r * (e^2_{t-1} + x_{t-1}P_{t-1}x_{t-1}') \qquad (4.1.3)'$$

4.1.2. A Transition Equation

The time-varying relationships between the stock returns and the explanatory variables implies that the state vector β_t changes over time. Let's assume that β_t is generated by a first-order Markov process,

$$\beta_t = M_t\beta_{t-1} + c_t + R_t u_t, \quad t = 1,2,\cdots,T \tag{4.1.5a}$$

where M_t is an mxm matrix, c_t is an mx1 vector, R_t is an mxg matrix and u_t is a gx1 vector of serially uncorrelated disturbances with mean zero and covariance matrix Q_t. Assume that the covariance matrix of u_t conditional on the past covariances show an ARCH(1) process as follows

$$u_t = Q_t^{1/2} w_t \tag{4.1.5b}$$

$$\{w_t, \ -\infty < t < \infty\} \sim \text{NID } (0,\ 1) \tag{4.1.5c}$$

$$Q_t = C + D \ .*\ u_{t-1}u_{t-1}' \tag{4.1.5d}$$

where C and D are the mxm symmetric matrices, respectively[1], and the notation of .* stands for element by element multiplication of two matrices. Consider a simple case where the M_t matrix remains unchanged over time, c_t is ignored, and u_t is seemingly uncorrelated with each other so that R_t is an identity matrix. Eq (4.1.2.a) will be substituted with the

---

1 $$Q_t = \begin{bmatrix} Q_{11t} & Q_{12t} & \cdots & Q_{1mt} \\ Q_{21t} & Q_{22t} & \cdots & Q_{2mt} \\ \cdots & \cdots & \cdots & \cdots \\ Q_{m1t} & Q_{m2t} & \cdots & Q_{mmt} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix} + \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & d_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ d_{m1} & d_{m2} & \cdots & d_{mm} \end{bmatrix} .*\ u_{t-1}u_{t-1}'$$

where $Q_{ijt} = \text{Cov}(u_{it},\ u_{jt})$
$= c_{ij} + d_{ij}{*}u_{i,t-1}u_{j,t-1}$
$= Q_{jit}, \quad \text{for } i,j = 1,2,\cdots,m$

following expression :

$$\beta_t = M\beta_{t-1} + u_t, \quad t = 1,2,\cdots,T \qquad (4.1.5a)'$$

As is in the previous case, if u_{t-1} is fixed and known at time t, then the distribution of u_t conditional on past observations will be normal with mean zero and variance-covariance matrix Q_t. Otherwise, it is not, in general, normally distributed. But for simplicity, it will be assumed to be normally distributed with mean zero and variance $E_{t-1}[u_t u_t']$.

$$E_{t-1}[u_t u_t'] = C + D .* E_{t-1}[u_{t-1}u_{t-1}'] \qquad (4.1.6a)$$

$$E_{t-1}[u^i{}_t u_{jt}'] = c_{ij} + d_{ij}\; E_{t-1}[u_{i,t-1}u_{j,t-1}'] \qquad (4.1.6b)$$

$$u_{t-1} = \beta_{t-1} - M\beta_{t-2} \quad \text{from the Eq (4.1.5a)'}$$

$$= (\beta_{t-1} - b_{t-1}) + (b_{t-1} - Mb_{t-2|t-1}) + M(b_{t-2|t-1} - \beta_{t-2})$$

where $b_{t-2|t-1}$ implies the optimal estimator of β_{t-2} obtained using the information available at time t-1. The conditional variance of u_{t-1} is given by

$$E_{t-1}[u_{t-1}u_{t-1}'] = P_{t-1} + v_{t-1}v_{t-1}' - P_{t-2,t-1|t-1}M' - MP_{t-2,t-1|t-1} + MP_{t-2|t-1}M' \qquad (4.1.7)$$

where $v_{t-1} = b_{t-1} - Mb_{t-2|t-1}$

$P_{t-2,t-1|t-1} = E_{t-1}[(b_{t-2|t-1} - \beta_{t-2})(b_{t-1} - \beta_{t-1})']$ and

$P_{t-2|t-1} = E_{t-1}[(b_{t-2|t-1} - \beta_{t-2})(b_{t-2|t-1} - \beta_{t-2})']$.

Substituting the Eq (4.1.7) into the Eq (4.1.6) produces the variance of u_t conditional on Ω_{t-1} as below,

$$E_{t-1}[u_t u_t'] = C + D .* [P_{t-1} + v_{t-1}v_{t-1}' - P_{t-2,t-1|t-1}M' - MP_{t-2,t-1|t-1} + MP_{t-2|t-1}M'] \qquad (4.1.7)'$$

From now on, let $h^c{}_t$ and $Q^c{}_t$ denote $E_{t-1}[\epsilon^2{}_t]$ and $E_{t-1}[u_t u_t']$ for notational convenience, respectively.

As can be seen in the Eq (4.1.6), the ij-th element of the variance-covariance matrix of u_t based on the observations including and up to time t-1 is linearly determined by the ij-th element of cross product of u_{t-1} conditional on the same information set. This model will be called as a matrix autoregressive conditional heteroscedasticity (MARCH). This can be regarded as a more general form of the standard ARCH model, in which only the variance terms of disturbances are assumed to follow ARCH processes, since the MARCH model enables the covariance terms to be utilized in the procedures of estimation of unknown parameters. Therefore we can get better estimators than otherwise since the covariance terms, which will be ignored in a standard ARCH model, carry some additional inforamtion. It follows that the difference, Q_t^c(ARCH) - Q_t^c(MARCH), will be positive definite.

Finally, it should be stressed that although the ϵ_t's are not mutually independent of each other, they are serially uncorrelated, so are the u_t's even if, like the ϵ_t's, they are also not independent of each other.

In order to derive the Kalman filter, we still need more assumptions on the initial state vector and the relations between the disturbance terms in both the measurement and transition equations and the initial state vector. First, the initial state vector, β_0, has a mean of b_0 and a covariance of

P_0, that is,

$$E(\beta_0) = b_0 \quad \text{and} \quad Var(\beta_0) = P_0. \tag{4.1.8a}$$

Second, the disturbances u_t and ϵ_t are uncorrelated with each other in all time periods, and uncorrelated with the initial state, that is,

$$E(u_t\epsilon_s') = 0 \qquad \text{for all } t,s = 1,2,\cdots,T$$

$$\text{and } E(u_t\beta_0') = 0, \quad E(\epsilon_t\beta_0') = 0 \qquad \text{for } t = 1,2,\cdots,T \tag{4.1.8b}$$

4.2 Derivation of the Kalman Filter

The Kalman filter provides a good way of computing the optimal estimates of the state vector, β_t, by applying a recursive procedure to the state space form, based on the information available at time t.

For the time being, let us assume that ϵ_t and u_t are normally distributed with means and variances described in Eq (4.1.1) and Eq (4.1.2). The Kalman filter consists of two crucial equation groups, the prediction equations and the updating equations. Let $b_{t|t-1}$ denote the mean of the distribution of β_t conditional on the information available at time t-1 and $P_{t|t-1}$ the mxm covariance matrix of the estimate of β conditional on the information available on time t-1. Given the assumption of normality of ϵ_t and u_t as well as the information available on time t-1, β_t is normally distributed with a mean of $b_{t|t-1}$ and a covariance matrix of $P_{t|t-1}$. Taking the conditional expectation on the both sides of Eq (4.1.2a)', we will obtain

$$b_{t|t-1} = Mb_{t-1}, \qquad t = 1,2,\cdots,T \tag{4.2.1a}$$

where b_{t-1} is assumed to be known at time t-1. From the Eq (4.1.5a)' and the Eq (4.2.1a),

$$\beta_t - b_{t|t-1} = M(\beta_{t-1} - b_{t-1}) + u_t$$

So, the variance of β_t conditional on Ω_{t-1} is given by

$$P_{t|t-1} = E_{t-1}[(\beta_t - b_{t|t-1})(\beta_t - b_{t|t-1})']$$

$$= MP_{t-1}M' + Q^c{}_t, \qquad t = 1,2,\cdots,T \tag{4.2.1b}$$

where $P_{t-1} = E[(\beta_{t-1} - b_{t-1})(\beta_{t-1} - b_{t-1})']$ is assumed to be known at time t. The above two equations, Eq (4.2.2a) and Eq (4.2.2b), are called as the prediction equations.

As the information about time t becomes available, the estimate of β can be updated. It should be noted that $\Omega_t = \{y_t, y_{t-1}, \cdots, y_1\}$ so that the distribution of β_t conditional on Ω_t is identical to that conditional on y_t. In order to get the optimal estimator of β_t based on y_t, we first need to find out the multivariate joint distribution of β_t and y_t. Under the assumption that ϵ_t and u_t conditional on Ω_{t-1} are normally distributed, β_t conditional on Ω_{t-1} is normally distributed with mean $b_{t|t-1}$ and covariance matrix $P_{t|t-1}$. The normality assumption also leads $y_t = x_t\beta_t + \epsilon_t$ conditional on Ω_{t-1} to be normally distributed with mean and variance as below.

$$E_{t-1}[y_t] = E_{t-1}[x_t\beta_t + u_t]$$

$$= x_t b_{t|t-1} \tag{4.2.2a}$$

The prediction error s_t based on Ω_{t-1} is given by

$$s_t = y_t - x_t b_{t|t-1}$$

$= x_t M[\beta_{t-1} - b_{t-1}] + \epsilon_t + x_t u_t$

So, the conditional variance of y_t is defined as

$$\begin{aligned} Var_{t-1}[y_t] &= E_{t-1}[s_t s_t'] \\ &= x_t[MP_{t-1}M' + Q^c_t]x_t' + h^c_t \\ &= x_t P_{t|t-1} x_t' + h^c_t \end{aligned} \qquad (4.2.2b)$$

The covariance matrix between β_t and y_t conditional on Ω_{t-1} is

$$E_{t-1}[(\beta_{t-1} - b_{t|t-1})s_t'] = P_{t|t-1}x_t' \qquad (4.2.2c)$$

From the Eq (4.2.1) and the Eq (4.2.2), we can deduce that the joint distribution of β_t and y_t conditional on Ω_{t-1} is normal with mean and covariance matrix as below.

$$\begin{bmatrix} \beta_t \\ y_t \end{bmatrix} \sim N\left[\begin{pmatrix} b_{t|t-1} \\ x_t b_{t|t-1} \end{pmatrix}, \begin{pmatrix} P_{t|t-1} & P_{t|t-1}x_t' \\ x_t P_{t|t-1} & x_t P_{t|t-1} x_t' + h_t^c \end{pmatrix}\right] \qquad (4.2.2d)$$

Applying the well known lemma[2] in statistics to Eq (4.2.2) produces the optimal estimator b_t of β_t and its covariance matrix P_t conditional on Ω_t as below.

$$b_t = b_{t|t-1} + P_{t|t-1}x_t' f_t^{-1}(y_t - x_t b_{t|t-1}) \qquad (4.2.3a)$$

and $P_t = P_{t|t-1} - P_{t|t-1}x_t' f_t^{-1} x_t P_{t|t-1}$ (4.2.3b)

where $f_t = x_t P_{t|t-1} x_t' + h^c_t, \qquad t = 1,2,\cdots,T$ (4.2.3c)

The Eq (4.2.3a), the Eq (4.2.3b), and the Eq(4.2.3c) are

---

[2] Lemma : Suppose that the pair of vectors x and y is jointly multivariate normal such that

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}\right]$$

Then the distribution of x conditional on y has a multinormal distribution with a mean of $\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$ and a covariance matrix of $\Sigma_{xx|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$.

called as the updating equations.

Taking the prediction equations and the updating equations together makes a unified system of equations, called as the Riccati Equations.

$$b_{t+1|t} = M[b_{t|t-1} + P_{t|t-1}x_t'f_t^{-1}(y_t - x_tb_{t|t-1})] \qquad (4.2.4a)$$

and $$P_{t+1|t} = M(P_{t|t-1} - P_{t|t-1}x_t'f_t^{-1}x_tP_{t|t-1})M' + Q^c_{t+1} \qquad (4.2.4b)$$

Using the Riccati equations, we can update the estimator of β_t as new information about observations is added to the information set.

On the other hand, the evaluation of the prediction, updating, and Riccati equations requires both $P_{t-2,t-1|t-1}$ and $P_{t-2|t-1}$ to be computed. As can be seen above, the covariance between β_{t-1} and β_{t-2} conditional on the information available at time t-1 is given by

$$P_{t-2,t-1|t-1} = E_{t-1}[(b_{t-2|t-1} - \beta_{t-2})(b_{t-1} - \beta_{t-1})']$$

$$= (x_{t-2}'x_{t-2})^{-1}x_{t-2}'e_{t-2}e_{t-1}'x_{t-1}(x_{t-1}'x_{t-1})^{-1} \qquad (4.2.5a)$$

-proof-

$$y_{t-1} = x_{t-1}\beta_{t-1} + \epsilon_{t-1} \qquad \text{from the Eq (4.1.1a)} \qquad (4.2.5b)$$

$$E_{t-1}(y_{t-1}) = x_{t-1}b_{t-1} \qquad (4.2.5c)$$

by taking the conditional expectation. Subtracting the Eq (4.2.5c) from the Eq (4.2.5b),

$$y_{t-1} - E_{t-1}(y_{t-1}) = x_{t-1}(\beta_{t-1} - b_{t-1}) + \epsilon_{t-1} \qquad (4.2.5d)$$

Multiplying both sides of the Eq (4.2.5d) by x_{t-1}' and then $(x_{t-1}'x_{t-1})^{-1}$,

$$\beta_{t-1} - b_{t-1} = (x_{t-1}'x_{t-1})^{-1}x_{t-1}'[y_{t-1} - E_{t-1}(y_{t-1}) - \epsilon_{t-1}] \qquad (4.2.5e)$$

In the similar way, we can get

$$\beta_{t-2} - b_{t-2|t-1} = (x_{t-2}'x_{t-2})^{-1}x_{t-2}'[y_{t-2} - E_{t-1}(y_{t-2}) - \epsilon_{t-2}] \quad (4.2.5f)$$

Taking the conditional expectation on the product of the Eq (4.2.5e) and the Eq (4.2.5f) gives the result described at the Eq (4.2.5a) since $E_{t-1}(\epsilon_{t-2}e_{t-1}') = E_{t-1}(\epsilon_{t-2}e_{t-1}') = 0$. -Q.E.D-

Next, it follows from the smoothing that the variance-covariance matrix of β_{t-2} conditional on the information available at time t-1 is given by

$$P_{t-2|t-1} = P_{t-2} + P_{t-2}^*(P_{t-1} - P_{t-1|t-2})P_{t-2}^{*\prime} \quad (4.2.5g)$$

where $P_{t-2}^* = P_{t-2}M'P_{t-1|t-2}^{-1}$.

On the other hand, smoothing algorithm provides a recursive procedure which yields the estimates of coefficients of the past periods using a cumulative set of the information which is already known to analysts. The fixed-interval smoothing estimates, proposed by Ansley and Kohn (1982), are given by

$$b_{t|T} = b_t + P_t^*(b_{t+1|T} - Mb_t) \quad (4.2.5h)$$

and $$P_{t|T} = P_t + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^{*\prime} \quad (4.2.5i)$$

where $P_t^* = P_tM'P_{t+1|t}^{-1}$, $t = T-1, T-2, \ldots, 1$ (4.2.5j)

4.3 Noteworthy Features of the Estimators Obtained Using the Kalman Filter

Under the assumption that ϵ_t and u_t conditional on Ω_{t-1} are normally distributed, the Kalman filtering leads the estimators of unknown parameters to have some desirable properties. (i) The Kalman filter produces the mean b_t of the conditional distribution of β_t as an optimal estimate of β_t in

the sense that it minimizes the mean squared error (Known as the minimum mean square estimator) :

$$b_t = E_t(\beta_t) = E(\beta_t | y_t) \tag{4.3.1}$$

As can be seen in the Eq (4.2.3a), the estimator b_t relies on the unknown parameters to be estimated in the next section and the mean and covariance of the initial state vector β_0. Furthermore, the parameter estimators obtained from minimizing the sum of squared prediction error, i.e., $\Sigma_t s_t^2$, are equivalent to those obtained from maximizing the conditional log-likelihood function provided that the Kalman filter converges. The next question then is when the Kalman filter converges to a steady state exponentially fast. When the transition equation is time invariant, the model should be observable and controllable as well as the characteristic roots of the transition matrix should be less than one in absolute value. When both measurement and transition equations are time invariant, the model should be detectable and stabilizable.

The final question is what the controllability, observability, detectability, and stabilizability of the model imply. Consider a simple model such that

$$y_t = x\beta_t, \quad \beta_t = M\beta_{t-1} + u_t$$

Then the model is controllable if $\text{Rank}[I, M, \cdots, M^{\tau-1}] = \tau$, observable if $\text{Rank}[x', M'x', \cdots, (M')^{\tau-1}x'] = \tau$, stabilizable if $\exists\ W_{mxn}$ s.t. $|\lambda_i(M + W')| < 1$ for $i = 1,\cdots,m$, and detectable if $\exists\ \Psi_{mxn}$ s.t. $|\lambda_i(M - \Psi x)| < 1$ for $i = 1,\cdots,m$.

So if the model satisfies the above conditions, the estimators of the unknown parameters obtained from the minimization of the sum of squared prediction errors will be those obtained from maximizing the conditional log-likelihood. It follows that b_t based on the MLE estimators of unknown parameters is equivalent to those based the MMSE estimators.

(ii) The estimator, b_t, is unconditionally unbiased since

$$\begin{aligned} b_t - E[b_t] &= b_t - E[E_t(\beta_t)] \\ &= b_t - E(\beta_t) \\ &= 0 \end{aligned}$$

(iii) The Kalman filter also yields the unconditional covariance matrix of the estimation error,

$$P_t = E_t[(\beta_t - E_t(\beta_t))(\beta_t - E_t(\beta_t))'] \qquad (4.3.2)$$

(iv) The conditional mean of y_t based on the information available at time t-1, $y_{t|t-1} = X_t b_{t|t-1}$, will be the MMSE of y_t.

(v) The distribution of the prediction error, $s_t = y_t - y_{t|t-1} = x_t M(\beta_{t-1} - b_{t-1}) + \epsilon_t + x_t u_t$, conditional on Ω_{t-1} is normal with mean and covariance matrix f_t as below and it also is independent of each other.

$$\begin{aligned} E_{t-1}(s_t) &= x_t M E_{t-1}(\beta_{t-1} - b_{t-1}) + E_{t-1}(\epsilon_t) + x_t E_{t-1}(u_t) \\ &= 0 \end{aligned}$$

$$Var_{t-1}(s_t) = x_t P_{t|t-1} x_t' + h^c{}_t = f_t$$

In other words,

$$s_t | \Omega_{t-1} \sim NID(0, f_t) \qquad (4.3.3)$$

Next, let us speculate a non-Gaussian model in which normality assumption on the disturbance terms is dropped. (i)

However, it still produces an optimal estimator in the sense that it minimizes the MSE within the class of all linear estimates (known as the minimum mean square linear estimator), even though the optimal estimate of β_t may not be the conditional mean of the state vector. (ii) b_t is also unconditionally unbiased. (iii) The unconditional covariance matrix of the estimation error is the P_t given by the Kalman filter. (iv) The conditional mean of y_t at time t-1, $y_{t|t-1}$, will be the MMSLE rather than the MMSE of y_t. (v) The prediction error, s_t, is distributed with mean zero and covariance matrix, f_t, and uncorrelated with that in different time periods.

4.4 Estimation of the Model

4.4.1. Constructing the Objective Function

In spite of some desirable properties which a maximum likelihood estimators (MLE) have, I will estimate our model by a non-linear least square method (NLLS) because of following reasons. First, the MLE method involves high computational costs than the NLLS method. Second, parameter estimators obtained from the NLLS method are equivalent to those obtained from the MLE method if the Kalman filter converges.

The next thing to do is to construct the objective function. In our model, the objective function to be minimized will be

the sum of the squared prediction errors, i.e.,

$$\begin{aligned} SSE_T(\theta) &= \sum_{t=1}^{T} (y_t - y_{t|t-1})^2 \\ &= \sum_{t=1}^{T} (y_t - x_t b_{t|t-1})^2 \end{aligned} \qquad (4.4.1a)$$

For testing, we need one additional assumption on the distribution of disturbance term conditional on the information available at time t-1. In fact, it is not likely that the distributions of disturbance terms in the state space model are independent and normal without precise knowledge of past disturbances. So it will cause the procedure of the hypothesis testing to be very much complicated. In order to avoid such a problem as above, we will assume that the disturbance terms conditional on the information available at time t-1 are normally distributed. It follows that the prediction error s_t is independently and normally distributed.

Suppose that it is known that the initial state vector, β_0, has a proper prior distribution with known mean, b_0, and bounded covariance matrix, P_0. Then the Kalman filter produces the exact likelihood function of the observations, y. In practice, proper prior distribution is hardly available. In this case, the initial distribution of β_0 must be specified in terms of a diffuse or non-informative prior.[3] In this

[3] Several algorithms have been proposed for obtaining appropriate prior distribution. Rosenberger (1973) and Wecker and Ansley(1983) proposed algorithms in which the elements of the initial state vector are treated as fixed so that they need to be estimated with either the maximum likelihood estimation technique or the GLS technique. Harvey and Phillips (1979), Burridge and Wallis (1985),

paper, however, I am going to use the OLS estimates and their covariance matrix, obtained from the first 60 observations, as the conditional mean and covariance matrix of the initial state vector, respectively.

4.4.2. Derivatives of the Objective Function

Taking the partial derivative of the sum of the squared prediction errors stated in (4.4.1a) with respect to the ith element of unknown parameters, θ,

$$\frac{\partial SSE_T(\theta)}{\partial \theta_i} = 2\sum_{t=1}^{T} s_t \frac{\partial s_t}{\partial \theta_i} \tag{4.4.2a}$$

where

$$\frac{\partial s_t}{\partial \theta_i} = - x_t \frac{\partial b_{t|t-1}}{\partial \theta_i} \quad .$$

To evaluate the Eq (4.4.2a), we need to compute the following partial derivatives:

$$\frac{\partial b_{t|t-1}}{\partial \theta_i} = \frac{\partial M}{\partial \theta_i} b_{t-1} + M \frac{\partial b_{t-1}}{\partial \theta_i} \tag{4.4.2b}$$

---

and Burmeister et al (1986) proposed the so-called 'big k' method in which the Kalman filter is initialized with a state covariance matrix of the form kI. The big k can reflect a high degree of uncertainty about the initial state, but it is numerically inaccurate. Ansley and Kohn (1985) proposed a general algorithm which employs data transformation to eliminate the dependence on the initial conditions. de Jong (1988,1989) presumed that the initial state vector are composed of two components, one with proper prior and the other with diffuse prior and developed an algorithm, which does not rely on the transformation of the data, by applying the Kalman fiter to the initial state vector. In his 1991 paper, de Jong developed the numerically stable algorithm for filtering, likelihood evaluation, generalized least squares computation and smoothing in which the diffuse Kalman filter is formulated in 'square root'. He argues that the algorithm increases numerical accuracy and stability.

$$\frac{\partial b_t}{\partial \theta_i} = \frac{\partial b_{t|t-1}}{\partial \theta_i} + \frac{\partial P_{t|t-1}}{\partial \theta_i} x_t' f_t^{-1} s_t - P_{t|t-1} x_t' f_t^{-2} \frac{\partial f_t}{\partial \theta_i} s_t + P_{t|t-1} x_t' f_t^{-1} \frac{\partial s_t}{\partial \theta_i} \tag{4.4.2c}$$

$$\frac{\partial P_{t|t-1}}{\partial \theta_i} = \frac{\partial M}{\partial \theta_i} P_{t-1} M' + M \frac{\partial P_{t-1}}{\partial \theta_i} M' + M P_{t-1} \frac{\partial M'}{\partial \theta_i} + \frac{\partial Q_t^c}{\partial \theta_i} \tag{4.4.2d}$$

$$\frac{\partial f_t}{\partial \theta_i} = -x_t \frac{\partial P_{t|t-1}}{\partial \theta_i} x_t' + \frac{\partial h_t^c}{\partial \theta_i} \tag{4.4.2e}$$

The evaluation of the Eq (4.4.2.d), in turn, requires the following partial derivatives:

$$\frac{\partial P_t}{\partial \theta_i} = \frac{\partial P_{t|t-1}}{\partial \theta_i} - \frac{\partial P_{t|t-1}}{\partial \theta_i} x_t' f_t^{-1} x_t P_{t|t-1} + P_{t|t-1} x_t' f_t^{-2} \frac{\partial f_t}{\partial \theta_i} x_t P_{t|t-1} - P_{t|t-1} x_t' f_t^{-1} x_t \frac{\partial P_{t|t-1}}{\partial \theta_i} \tag{4.4.2f}$$

$$\frac{\partial h_t^c}{\partial \theta_i} = \frac{\partial q}{\partial \theta_i} + \frac{\partial r}{\partial \theta_i} (e_{t-1}^2 + x_{t-1} P_{t-1} x_{t-1}') + r\,[2\, e_{t-1} \frac{\partial e_{t-1}}{\partial \theta_i} + x_{t-1} \frac{\partial P_{t-1}}{\partial \theta_i} x_{t-1}'] \tag{4.4.2g}$$

$$\begin{aligned} \frac{\partial Q_t^c}{\partial \theta_i} = {} & \frac{\partial C}{\partial \theta_i} + \frac{\partial D}{\partial \theta_i} .* (P_{t-1} + v_{t-1} v_{t-1}' - P_{t-2,t-1|t-1} M' - M P_{t-2,t-1|t-1} \\ & + M P_{t-2|t-1} M') + D .* [\frac{\partial P_{t-1}}{\partial \theta_i} + \frac{\partial v_{t-1}}{\partial \theta_i} v_{t-1}' \\ & + v_{t-1} \frac{\partial v_{t-1}'}{\partial \theta_i} - \frac{\partial P_{t-2,t-1|t-1}}{\partial \theta_i} M' - P_{t-2,t-1|t-1} \frac{\partial M'}{\partial \theta_i} \\ & - \frac{\partial M}{\partial \theta_i} P_{t-2,t-1|t-1} - M \frac{\partial P_{t-2,t-1|t-1}}{\partial \theta_i} + \frac{\partial M}{\partial \theta_i} P_{t-2|t-1} M' \\ & + M \frac{\partial P_{t-2|t-1}}{\partial \theta_i} M' + M P_{t-2|t-1} \frac{\partial M'}{\partial \theta_i}] \end{aligned} \tag{4.4.2h}$$

which can be obtained from the prediction equations, respectively. Putting the above equations, the Eq (4.4.2), all together will enable the gradient vector to be evaluated,

$$GRD(\theta) = \frac{\partial SSE_T(\theta)}{\partial \theta} \tag{4.4.2i}$$

4.4.3. Computing the Hessian Matrix

As can be seen above, this optimization program is non-linear so we have to rely on an iterative procedure for estimation of the unknown parameters. Most of iterative methods developed recently requires the evaluation of the Hessian matrix for iteration. The ith row and jth column element of the Hessian matrix can be obtained by differentiating (4.4.2a) with respect to the jth element of θ.

$$\frac{\partial^2 SSE_T(\theta)}{\partial\theta_i\partial\theta_j} = -2\sum_{t=1}^{T}\left[-x_t\frac{\partial b_{t|t-1}}{\partial\theta_i}x_t\frac{\partial b_{t|t-1}}{\partial\theta_j} + (y_t - x_t b_{t|t-1})\, x_t\frac{\partial^2 b_{t|t-1}}{\partial\theta_i\partial\theta_j}\right] \tag{4.4.3a}$$

In order to evaluate the Eq (4.4.3a), we need the following calculations,

$$\frac{\partial^2 b_{t|t-1}}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 M}{\partial\theta_i\partial\theta_j}b_{t-1} + \frac{\partial M}{\partial\theta_i}\frac{\partial b_{t-1}}{\partial\theta_j} + \frac{\partial M}{\partial\theta_j}\frac{\partial b_{t-1}}{\partial\theta_i} + M\frac{\partial^2 b_{t-1}}{\partial\theta_i\partial\theta} \tag{4.4.3b}$$

$$\begin{aligned}\frac{\partial^2 b_t}{\partial\theta_i\partial\theta_j} &= \frac{\partial^2 b_{t|t-1}}{\partial\theta_i\partial\theta_j} + \frac{\partial^2 P_{t|t-1}}{\partial\theta_i\partial\theta_j}\frac{x_t s_t}{f_t} - \frac{\partial P_{t|t-1}}{\partial\theta_i}\frac{x_t' s_t}{f_t^2} + \frac{\partial P_{t|t-1}}{\partial\theta_i}\frac{x_t'}{f_t}\frac{\partial s_t}{\partial\theta_j}\\ &- \frac{\partial P_{t|t-1}}{\partial\theta_j}\frac{x_t}{f_t^2}\frac{\partial f_t}{\partial\theta_i}s_t + 2P_{t|t-1}\frac{x_t'}{f_t^3}\frac{\partial f_t}{\partial\theta_j}\frac{\partial f_t}{\partial\theta_i}s_t - P_{t|t-1}\frac{x_t'}{f_t^2}\frac{\partial^2 f_t}{\partial\theta_i\partial\theta_j}s_t\\ &- P_{t|t-1}\frac{x_t'}{f_t^2}\frac{\partial f_t}{\partial\theta_i}\frac{\partial s_t}{\partial\theta_j} + \frac{\partial P_{t|t-1}}{\partial\theta_j}\frac{x_t'}{f_t}\frac{\partial s_t}{\partial\theta_i} - P_{t|t-1}\frac{x_t'}{f_t^2}\frac{\partial f_t}{\partial\theta_j}\frac{\partial s_t}{\partial\theta_j}\\ &+ P_{t|t-1}\frac{x_t'}{f_t}\frac{\partial^2 s_t}{\partial\theta_i\partial\theta_j}\end{aligned} \tag{4.4.3c}$$

where

$$\frac{\partial^2 s_t}{\partial\theta_i\partial\theta_j} = -\,x_t\frac{\partial^2 b_{t|t-1}}{\partial\theta_i\partial\theta_j} \tag{4.4.3d}$$

$$\frac{\partial^2 P_{t|t-1}}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 M}{\partial\theta_i\partial\theta_j} P_{t-1} M' + \frac{\partial M}{\partial\theta_i}\frac{\partial P_{t-1}}{\partial\theta_j} M' + \frac{\partial M}{\partial\theta_i} P_{t-1}\frac{\partial M'}{\partial\theta_j}$$
$$+ \frac{\partial M}{\partial\theta_j}\frac{\partial P_{t-1}}{\partial\theta_i} M' + M\frac{\partial^2 P_{t-1}}{\partial\theta_i\partial\theta_j} M' + M\frac{\partial P_{t-1}}{\partial\theta_i}\frac{\partial M'}{\partial\theta_j}$$
$$+ \frac{\partial M}{\partial\theta_j} P_{t-1}\frac{\partial M'}{\partial\theta_i} + M\frac{\partial P_{t-1}}{\partial\theta_j}\frac{\partial M'}{\partial\theta_i} + M P_{t-1}\frac{\partial^2 M'}{\partial\theta_i\partial\theta_j} \tag{4.4.3e}$$
$$+ \frac{\partial^2 Q_t^c}{\partial\theta_j\partial\theta_j}$$

$$\frac{\partial^2 P_t}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 P_{t|t-1}}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 P_{t|t-1}}{\partial\theta_i\partial\theta_j}\frac{x_t' x_t}{f_t} P_{t|t-1} + \frac{\partial P_{t|t-1}}{\partial\theta_i}\frac{x_t'}{f_t^2}\frac{\partial f_t}{\partial\theta_j} x_t P_{t|t-1}$$
$$- \frac{\partial P_{t|t-1}}{\partial\theta_i}\frac{x_t' x_t}{f_t}\frac{\partial P_{t|t-1}}{\partial\theta_j} + \frac{\partial P_{t|t-1}}{\partial\theta_j}\frac{x_t'}{f_t^2}\frac{\partial f_t}{\partial\theta_i} x_t P_{t|t-1} - 2 P_{t|t-1}\frac{x_t'}{f_t^3}\frac{\partial f_t}{\partial\theta_j}\frac{\partial f_t}{\partial\theta_i} x_t P_{t|t-1} \tag{4.4.3f}$$
$$+ P_{t|t-1}\frac{x_t'}{f_t^2}\frac{\partial^2 f_t}{\partial\theta_i\partial\theta_j} x_t P_{t|t-1} + P_{t|t-1}\frac{x_t'}{f_t^2}\frac{\partial f_t}{\partial\theta_i} x_t\frac{\partial P_{t|t-1}}{\partial\theta_j} - \frac{\partial P_{t|t-1}}{\partial\theta_j}\frac{x_t' x_t}{f_t}\frac{\partial P_{t|t-1}}{\partial\theta_i}$$
$$+ P_{t|t-1}\frac{x_t'}{f_t^2}\frac{\partial f_t}{\partial\theta_j} x_t\frac{\partial P_{t|t-1}}{\partial\theta_i} - P_{t|t-1}\frac{x_t' x_t}{f_t}\frac{\partial^2 P_{t|t-1}}{\partial\theta_i\partial\theta_j}$$

$$\frac{\partial^2 f_t}{\partial\theta_i\partial\theta_j} = -x_t\frac{\partial^2 P_{t|t-1}}{\partial\theta_i\partial\theta_j} x_t' + \frac{\partial^2 h_t^c}{\partial\theta_i\partial\theta_j} \tag{4.4.3g}$$

$$\frac{\partial^2 h_t^c}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 q}{\partial\theta_i\partial\theta_j} + \frac{\partial^2 r}{\partial\theta_i\partial_j}\left(e_{t-1}^2 + x_{t-1} P_{t-1} x_{t-1}'\right)$$
$$+ \frac{\partial r}{\partial\theta_i}\left(2 e_{t-1}\frac{\partial e_{t-1}}{\partial\theta_j} + x_{t-1}\frac{\partial P_{t-1}}{\partial\theta_j} x_{t-1}'\right)$$
$$+ \frac{\partial r}{\partial\theta_j}\left(2 e_{t-1}\frac{\partial e_{t-1}}{\partial\theta_i} + x_{t-1}\frac{\partial P_{t-1}}{\partial\theta_i} x_{t-1}'\right) \tag{4.4.3h}$$
$$+ r\left(2\frac{\partial e_{t-1}}{\partial\theta_j}\frac{\partial e_{t-1}}{\partial\theta_i} + 2 e_{t-1}\frac{\partial^2 e_{t-1}}{\partial\theta_i\partial_j} + x_{t-1}\frac{\partial^2 P_{t-1}}{\partial\theta_i\partial\theta_j} x_{t-1}'\right)$$

$$
\begin{aligned}
\frac{\partial^2 Q_t^c}{\partial\theta_i\partial\theta_j} = {} & \frac{\partial^2 C}{\partial\theta_i\partial\theta_j} + \frac{\partial^2 D}{\partial\theta_i\partial\theta_j} .* (P_{t-1} + v_{t-1}v_{t-1}' - P_{t-2,t-1|t-1}M' - MP_{t-2,t-1|t-1} \\
& + MP_{t-2|t-1}M') + \frac{\partial D}{\partial\theta_i} .* [\frac{\partial P_{t-1}}{\partial\theta_j} + \frac{\partial v_{t-1}}{\partial\theta_j} v_{t-1}' + v_{t-1}\frac{\partial v_{t-1}'}{\partial\theta_j} \\
& - \frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_j} M' - P_{t-2,t-1|t-1}\frac{\partial M'}{\partial\theta_j} - \frac{\partial M}{\partial\theta_j} P_{t-2,t-1|t-1} \\
& - M\frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_j} + \frac{\partial M}{\partial\theta_j} P_{t-2|t-1}M' + M\frac{\partial P_{t-2|t-1}}{\partial\theta_j} M' \\
& + MP_{t-2|t-1}\frac{\partial M'}{\partial\theta_j}] + \frac{\partial D}{\partial\theta_j} .* [\frac{\partial P_{t-1}}{\partial\theta_i} + \frac{\partial v_{t-1}}{\partial\theta_i} v_{t-1}' + v_{t-1}\frac{\partial v_{t-1}}{\partial\theta_i} \\
& - \frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_i} M' - P_{t-2,t-1|t-1}\frac{\partial M'}{\partial\theta_i} - \frac{\partial M}{\partial\theta_i} P_{t-2,t-1|t-1} - M\frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_i} \\
& + \frac{\partial M}{\partial\theta_i} P_{t-2|t-1}M' + MP_{t-2|t-1}\frac{\partial M'}{\partial\theta_i}] + D.*[\frac{\partial^2 P_{t-1}}{\partial\theta_i\partial\theta_j} \\
& + \frac{\partial^2 v_{t-1}}{\partial\theta_i\partial\theta_j} v_{t-1}' + \frac{\partial v_{t-1}}{\partial\theta_i}\frac{\partial v_{t-1}'}{\partial\theta_j} + \frac{\partial v_{t-1}}{\partial\theta_j}\frac{\partial v_{t-1}'}{\partial\theta_i} + v_{t-1}\frac{\partial^2 v_{t-1}'}{\partial\theta_i\partial\theta_j} \\
& - \frac{\partial^2 P_{t-2,t-1|t-1}}{\partial\theta_i\partial\theta_j} M' - \frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_i}\frac{\partial M'}{\partial\theta_j} - \frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_j}\frac{\partial M'}{\partial\theta_i} - P_{t-2,t-1|t-1}\frac{\partial^2 M'}{\partial\theta_i\partial\theta_j} \\
& - \frac{\partial^2 M}{\partial\theta_i\partial\theta_j} P_{t-2,t-1|t-1} - \frac{\partial M}{\partial\theta_i}\frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_j} - \frac{\partial M}{patial\theta_j}\frac{\partial P_{t-2,t-1|t-1}}{\partial\theta_i} - M\frac{\partial^2 P_{t-2,t-1}}{\partial\theta_i\partial\theta_{\_}} \\
& + \frac{\partial^2 M}{\partial\theta_i\partial\theta_j} P_{t-2|t-1}M' + \frac{\partial M}{\partial\theta_i}\frac{\partial P_{t-2|t-1}}{\partial\theta_j} M' + \frac{\partial M}{\partial\theta_i} P_{t-2|t-1}\frac{\partial M'}{\partial\theta_j} + \frac{\partial M}{\partial\theta_j}\frac{\partial P_{t-2|t-1}}{\partial\theta_j} M' \\
& + M\frac{\partial^2 P_{t-2|t-1}}{\partial\theta_i\partial\theta_j} M' + M\frac{\partial P_{t-2|t-1}}{\partial\theta_i}\frac{\partial M'}{\partial\theta_j} + \frac{\partial M}{\partial\theta_j} P_{t-2|t-1}\frac{\partial M'}{\partial\theta_i} + M\frac{\partial P_{t-2|t-1}}{\partial\theta_j}\frac{\partial M'}{\partial thetha_i} \\
& + MP_{t-2|t-1}\frac{\partial^2 M'}{\partial\theta_i\partial\theta_j}]
\end{aligned}
$$

(4.4.3i)

Combining all the partial derivatives given in the Eq (4.4.2) and the Eq (4.4.3) will enable the Hessian matrix to be computed.

$$HESS(\theta) = \frac{\partial^2 SSE_T(\theta)}{\partial\theta\partial\theta'} \qquad (4.4.3j)$$

From the Eq (4.4.2i) and the Eq (4.4.3j), the iteration algorithm, proposed by Davidon (1959) and Fletcher & Powell

(1963), for estimating the unknown parameters will be given by

$$\theta = \theta_0 - \lambda_0 HESS(\theta_0)^{-1} GRD(\theta_0)$$

where λ_0 is determined by a cubic interpolation of $SSE_T(\theta)$ along the current search direction.

CHAPTER 5
RESULTS OF EMPIRICAL WORK

5.1 Description of Data

In this section, I am going to describe the structure of the data set which is used in this paper. As can be seen in the Eq (4.1.2), the current stock return index of the U.S. stock markets are attempted to be explained by a linear combination of constant, current term spread (= long-term interest rate - short-term interest rate), current inflation rate based on the producer price index, future growth rates of the industrial production of the U.S.A.. All of the series are the same as the ones used in the 1990 Schwert paper and the 1990 Pagan & Schwert paper[1].

5.1.1. Data on the Monthly Stock Returns

For the period 1889 to 1925, the monthly stock return index is measured by the capital gain returns to the Dow Jones composite index (1972) plus the dividend yield from the value-weighted portfolio of NYSE stocks constructed by the Cowles Commission (1939, pp 168-169). For the period 1926 to 1987, it is equal to the monthly stock returns including dividends to the value-weighted portfolio of all NYSE stocks constructed

[1] I am very grateful to Dr. Adrian Pagan for sending me all the data sets.

by the Center for Research in Security Prices (CRSP).

5.1.2. Data on the Short-term Interest Rates

For the period 1889 to 1925, the short term interest rate is estimated using the four to six month commercial paper rates in New York from Macaulay (1938, Table 10, pp A141-A161) and the result of the simple linear regression of CRSP yields on Macaulay yields from 1926 to 1937. For the period 1926 to 1987, it is equal to the monthly tax-free yields on the U.S. Government security, which matures the end of the month, in the Government Bond File constructed by CRSP.

5.1.3. Data on the Inflation Rates

For 1889, the inflation rate is measured by the monthly inflation rate of the Warren and Pearson (1933) index of the U.S. producer prices. For the period 1890 to 1987, it is calculated from the not seasonally adjusted U.S. producer price index in the Bureau of Labor Statistics.

5.1.4. Data on the industrial Production

For the period 1889 to 1918, the growth rate of industrial production is estimated using the Babson's Index of the physical volume of business activity from Moore (1961, pp 130) and the average ratio of Babson to adjusted industrial production for 1919-1939. For the period 1919-1987, it is calculated from the index of industrial production of the

Federal Reserve Board (1986) and Citibase (1986).

5.1.5. Long-term Interest Rates and Term Spreads

For the period 1889 to 1918, the long-term interest rate is estimated from Macaulay's (1938, Table 10, pp. A141-A161) railroad bond yield index using the average ratio to the Moody's Aa bond yield. For the period 1919 to 1987, it implies the Moody's Aa bond yield.

The term spread is defined as the long-term interest rate minus the short-term interest rate.

5.2 Results of Estimation of the Model

5.2.1. Previous Results Based on the Constant Parameter Models

In this section, I am going to summarize the previous results associated with the relationships between the U.S. stock returns and the macroeconomic variables described above. It will be helpful for comparing our estimation results with the results from previous work on this topic.

First of all, large fraction of the variation in the stock returns can be accounted for by future real activities such as real GNP, industrial production, and investment (Fama, 1990). Especially, the future growth rates of industrial production among the real activities has a significant and positive effect on the current stock returns (Kaul & Seyhun, 1990; Shanken & Weinstein, 1990).

Second, during the post-war period, stock returns are

negatively related to inflation rate, to be more precise, relative price variability. The negative effect of relative price variability on the future output and thus on the contemporaneous stock returns are largely caused by the supply shocks in the 1970s (Kaul & Seyhun, 1990). In this paper, however, the inflation rate will be used as a proxy for the relative price variability.

Third, the term spread is positively correlated with expected stock returns because it carries some information on how high output growth is anticipated (Fama, 1990).

As will be seen in the latter section, however, relationships between stock returns and the three macroeconomic variables, found in the previous work, are valid only for a certain time period in this model.

5.2.2. Results of Estimation of the Model

Originally, the general model described in the Eq (4.1.2) and the Eq (4.1.5), was attempted to be estimated by the maximum likelihood estimation method. However, it involves very high computational costs. Instead, a simpler version of the time-varying parameter model which does not allow the disturbance terms of both the measurement and the transition equations to follow an ARCH process is estimated by the non-linear least square method (NLLS). That is,

$$y_t = x_t\beta_t + \epsilon_t \qquad (5.2.1)$$

$$\beta_t = M\beta_{t-1} + u_t$$

$\epsilon_t \sim N(0, H)$ and $u_t \sim N(0, Q)$

where H is a positive scalar, M is a 4 by 4 matrix and Q is a positive definite 4 by 4 symmetric matrix. The optimal estimates of M, Q and H are obtained from minimization of the sum of squared prediction errors. That is,

$$\min_{\theta} \sum_{t=1}^{T} (y_t - y_{t|t-1})^2 = \min_{\theta} \sum_{t=1}^{T} (y_t - x_t b_{t|t-1})^2 \qquad (5.2.2.)$$

I estimate the Eq (5.2.2.) by the iteration method, proposed by Davidon, Fletch and Powell, which is available in the Gauss program. The results of estimation of the 27 unknown parameters are summarized in Table 5.2.1. Unfortunately most of the parameter estimates appear to be statistically insignificantly different from zeros even at 10 % of significance level. But the optimization procedure in the Gauss program depends upon the numerical gradient vector and Hessian matrix of the objective function, which are obtained through lots of the approximation. This procedure may lead the estimates of the standard errors to be high so it may suppress the t-statistics of the estimates.

One of the several features of the parameter estimates is that each element of the state space vector ß is largely determined by its own lagged value rather than lagged values of other elements. And relatively high values of estimates of M23, M32 and M42 suggest that there exists some degree of correlation between the coefficient on the term spread, that on the inflation rate and that on the industrial production.

Higher variance estimate of the coefficient on the inflation rate may be the evidence that the relationship between the inflation rate and the stock return is more unstable than those between either the term spread or the industrial production and the stock returns.

Table 5.2.1 Results of Estimation of the Kalman Filter Model

Parameter	Estimate	S.E.	Ratio	Prob-value
M11	0.99477	0.57769	1.72198	0.08507
M12	0.00164	0.09432	0.01738	0.98613
M13	0.00026	0.57119	0.00045	0.99964
M14	-0.00016	0.57656	-0.00027	0.99979
M21	-0.00150	0.57769	-0.00259	0.99793
M22	0.44081	5.38708	0.08183	0.93478
M23	-0.10221	0.99648	-0.10257	0.91830
M24	0.01043	0.66989	0.01557	0.98758
M31	-0.00372	0.57769	-0.00644	0.99486
M32	-0.10443	0.51498	-0.20278	0.83931
M33	0.95322	0.57633	1.65396	0.09814
M34	-0.00440	0.57745	-0.00761	0.99392
M41	0.00266	0.57769	0.00461	0.99632
M42	0.10656	0.31374	0.33966	0.73411
M43	0.02767	0.57299	0.04829	0.96148
M44	0.98295	0.57687	1.70392	0.08840
Q11	0.04200	0.57769	0.07270	0.94204
Q21	-0.07036	0.57769	-0.12180	0.90306
Q22	0.05432	0.57769	0.09403	0.92509
Q31	-0.00011	0.57769	-0.00019	0.99985
Q32	0.01140	0.57769	0.01973	0.98426
Q33	0.46500	0.57769	0.80493	0.42086
Q41	0.00042	0.57769	0.00072	0.99942
Q42	-0.03069	0.57769	-0.05312	0.95764
Q43	-0.00239	0.57769	-0.00413	0.99670
Q44	0.01256	0.57769	0.02174	0.98266
H	0.04200	0.57769	0.07270	0.94204

Based on the parameter estimates shown in Table 5.2.1, I compute the estimates of β_t conditional on the information set available at time t, that is, the estimates defined as the Eq

(4.2.3a). Figure (5.1) shows the plot of the estimate of the constant term of the updating equation. The estimated series appears to be similar to that of the stock returns with smaller scale.

In his paper (1990), Fama argues that since a high term spread plays a role as a signal that there will be some improvement in output growth in the near future and expected stock return is positively correlated with the output growth, the term spread has a positive correlation with the stock returns. As can be seen in Figure (5.2), the path of the estimates of the coefficient on the term spreads shows high frequency over time. This implies that the information the term spreads carry is noisy. For example, during the Great Depression most of the firms in the U.S.A. probably should suffer from insufficient funds. So they probably issued corporate bonds with high yields in order to relieve the fund pressure rather than to do good projects in the recession. This high term spread occurred with the stock market crash almost at the same time periods. That is why during the Depression the term spreads seem to be highly negatively correlated with the stock returns. In addition, the fact that during the post-war period the term spread shows some features of non-stationary process itself may be able to explain why the coefficient estimates move around zero with high frequency. This may imply either that the relationship between the term spread and the stock return dynamically

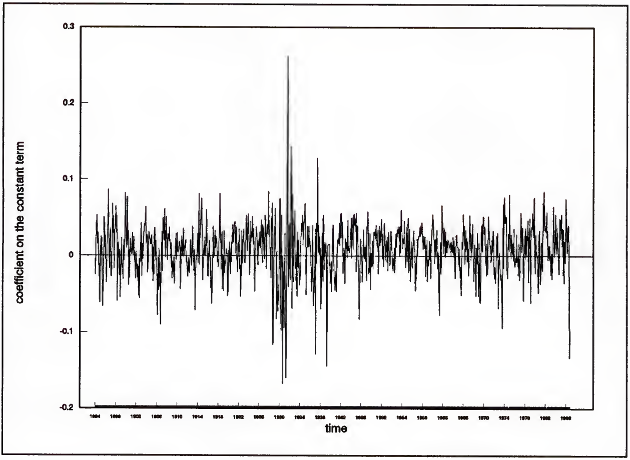

Figure 5.1 : The estimates of the constant term of the updating equations using the U.S. monthly stock return indexes of the CRSP

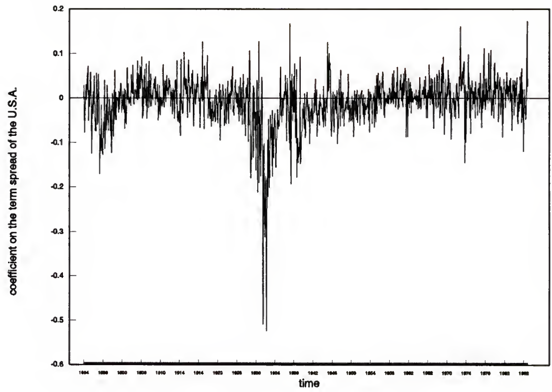

Figure 5.2 : The estimates of the coefficient on the term spread of the updating equations using the U.S. monthly stock return indexes of the CRSP

varies over time or that the term spreads have little explanatory power over the stock returns.

Compared with the term spread, the inflation rate seems to have a more stable relationship with the stock return. As can be seen in Figure (5.3), the path of the estimates of the coefficient on the inflation rate shows less frequency than that of the term spread. Next, the inflation caused by the demand shocks in general tends to have favorable effects on the firms' business because usually it accompanies economic expansion while that caused by the supply shocks tends to have unfavorable effects on the firms' business because it usually occurs together with economic recessions. For example, the supply-side inflation around the WW1, the 1970s and the early 1980s led the stock market down so during these periods the stock returns tends to be negatively related to the inflation. But the deflation around the Great Depression also partly led the stock market down and the inflation after the WW2 till the Korean war produced economic boom so during these periods the stock returns tends to be positively related to the inflation. Thus this partly supports Kaul and Seyhun (1990)'s argument that during the post-war period stock returns are negatively related to inflation rates since they focus on the effects of the supply-side inflation by the 'relative price variability'.

Until the early 1960s, as shown in Figure (5.4), the industrial production in general seems to be positively correlated with the stock returns even if there have been

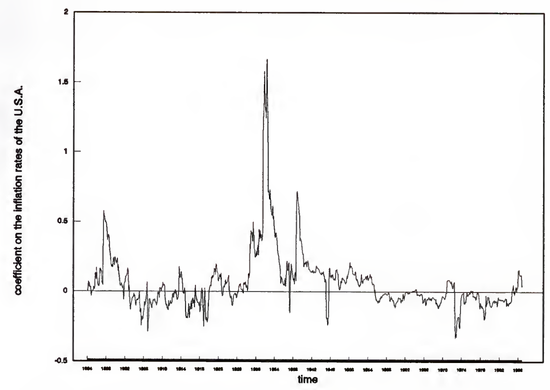

Figure 5.3 : The estimates of the coefficient on the inflation rate of the updating equations using the U.S. monthly stock return indexes of the CRSP

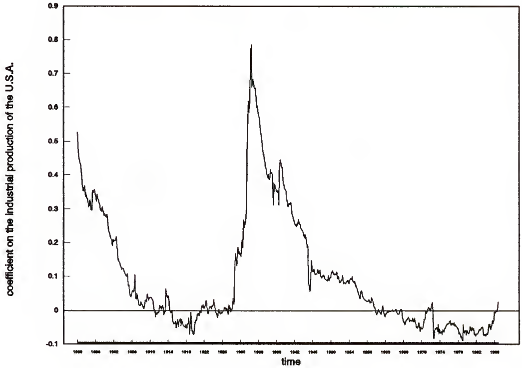

Figure 5.4 : The estimates of the coefficient on the industrial production of the updating equations using the U.S. monthly stock return indexes of the CRSP

found some variations in the scale of the estimates of the coefficient on it. Since the early 1960s the U.S. economy looks like to rapidly lose the strong incentives in the production sides but the stock markets meet kinds of rush as lots of funds concentrate on the financial markets for the purpose of speculation. That is why during that period the industrial production appears to be negatively related to the stock returns.

In summary, we should be very careful about the relationships between the stock returns and the macroeconomic varaibles. And there is little evidence that these relationships are stable or fixed over time.

The time-varying model employed in this thesis produces relatively good explanatory power over the variations in the stock returns. R^2 is 0.92785 from the updating equations and 0.12698 from the prediction equations. In Figure (5.5) and Figure (5.6), the forecasted stock returns obtained from the updating equations and the prediction equations are plotted with the actual stock returns, respectively.

On the other hand, I estimate the volatility of the stock returns by the squares of the prediction errors, and plot the results in Figure (6.1). This figure supports the results found in the previous literature that the volatility of stock returns varies over time and becomes high around economic recession, oil shocks, and banking panic.

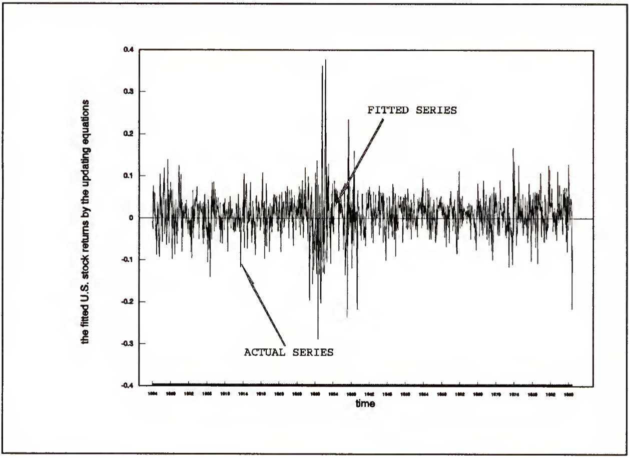

Figure 5.5 : The U.S. monthly stock returns and their fitted series by the updating equations using the U.S. monthly stock return indexes of the CRSP

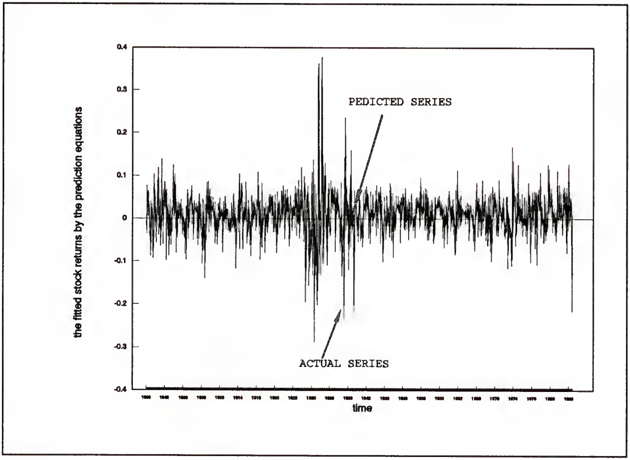

Figure 5.6 : The U.S. monthly returns and their forecasted series by the prediction equations using the U.S. monthly stock return indexes of the CRSP

5.2.3. Test on Parameter Constancy

In this section, the constancy of parameter β_t over time will be tested using the Likelihood Ratio statistic.

If the hypothesis that the coefficients β_t remain constant is true, then the M matrix in the transition equation must be an identity matrix and the variance of β_t must be equal to zero. As the result, the null and alternative hypotheses for testing the constancy of coefficients are given by

H_0 : $\beta_t = \beta_{t-1} = \cdots = \beta_1$, i.e., M is identity matrix, and β_t has zero variance, $t = 1, 2, \cdots, T$

H_a : Either that M is not an identity matrix or that there exists at least a non-zero variance of β_t.

The null hypothesis implies a simple linear model with constant coefficients, i.e., the estimate of β can be estimated by an OLS method.

The hypotheses described above can be expressed in a more general form as below.

$$H_0 : R\theta - \psi = 0$$

$$H_a : R\theta - \psi \neq 0 \qquad (5.2.3)$$

where R is the nxk known matrix of constants, θ is the kx1 vector of unknown parameters, and ψ is the nx1 vector.

In order to get a test statistic of the parameter constancy test, we need to estimate the model described in the previous section. This is done by maximising the following log-

likelihood function with respect to the unknowns:

$$LogL = -\frac{1}{2}TLn(2*\pi) - \frac{1}{2}\sum_{t=1}^{T} Ln(ff_t) - \frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - x_t b_t)^2}{ff_t} \quad (5.2.4)$$

where

$$ff_t = x_t * P_t * x_t' + H.$$

Estimating the model under the null, we utilize information on both y_t and x_t simultaneously. Thus for the parameter constancy test it is more reasonable to construct the log-likelihood function based on the updating equations rather than on the prediction equations. The MLE result is summarized in Table (5.2.2).

There have been proposed three useful misspecification tests which can apply to this kind of case : the likelihood ratio test (LR test), the Wald test (W test) and the Lagrange multiplier test (LM test). Let θ_0 and θ_a denote the restricted and unrestricted estimates of θ, respectively. First the LR test is based upon a comparison of the supremum of $L(\theta)$ under the null hypothesis with that under the alternative hypothesis. The LR test statistic is given by

$$LR = 2[L(\theta_a) - L(\theta_0)] \quad (5.2.5a)$$

Applying the Taylor expansion to $L(\theta_a)$ around θ_0 and ignoring the higher order terms than the second order, we obtain

$$L(\theta_a) \approx L(\theta_0) + (\theta_a - \theta_0)'H(\theta_a)(\theta_a - \theta_0)/2 \quad (5.2.5b)$$

where $H(\theta_a)$ is the hessian matrix evaluated at unrestricted estimates of θ. Rearranging the Eq (5.2.5b) will give

$$(\theta_a - \theta_0)'H(\theta_a)(\theta_a - \theta_0) \approx 2[L(\theta_a) - L(\theta_0)]$$

where under the null hypothesis $(\theta_a - \theta_0)'H(\theta_a)(\theta_a - \theta_0)$ is

known as being asymptotically χ^2-distributed with n degree of freedom so is $2[L(\theta_a) - L(\theta_0)]$.

Table 5.2.2 Results of Estimation of the Model under H_a

Parameter	Estimate	S.E.	Ratio	Prob-value
M11	0.99836	0.02000	49.92786	0.00000
M12	-4.68223	0.01998	-234.30924	0.00000
M13	0.40468	0.01934	20.92403	0.00000
M14	0.31499	0.01990	15.82763	0.00000
M21	-0.00144	0.01996	-0.07203	0.94258
M22	0.34642	0.01998	17.33508	0.00000
M23	-0.10078	0.01990	-5.06295	0.00000
M24	0.01257	0.01993	0.63067	0.52826
M31	-0.00388	0.01987	-0.19546	0.84503
M32	0.00075	0.01990	0.03791	0.96976
M33	0.92057	0.01990	46.25440	0.00000
M34	-0.01991	0.01993	-0.99880	0.31789
M41	0.00275	0.01993	0.13796	0.89028
M42	-0.07199	0.01989	-3.61879	0.00030
M43	0.05017	0.01931	2.59855	0.00936
M44	0.94769	0.01817	52.16018	0.00000
Q11	0.04200	0.01977	2.12486	0.03360
Q21	0.14533	0.01996	7.28045	0.00000
Q22	0.05471	0.01996	2.74096	0.00613
Q31	-0.01681	0.02001	-0.83977	0.40104
Q32	0.01170	0.01988	0.58844	0.55624
Q33	0.41565	0.01999	20.79626	0.00000
Q41	-0.01301	0.01993	-0.65283	0.51387
Q42	-0.03717	0.02001	-1.85784	0.06319
Q43	-0.06195	0.02001	-3.09608	0.00196
Q44	0.34269	0.01985	17.26082	0.00000
H	0.00003	0.00000	8.65876	0.00000

The LR test requires relatively high computational cost since the model must be estimated under both the null and alternative hypotheses.

Given the fact that θ_a is consistent for θ and, under H_0, $R\theta - \psi = 0$ so that $R\theta - \psi$ tends to be a null vector, the W test is a test of the joint significance of the elements of

$R\theta_a - \psi$. Using the fact that $R\theta_a - \psi$ is asymptotically normally distributed with mean zero and variance $R\Lambda^{-1}R'$ where Λ is the information matrix evaluated at unrestricted estimates of θ, we can obtain

$$W = (R\theta_a - \psi)'(R\Lambda^{-1}(\theta_a)R')^{-1}(R\theta_a - \psi) \sim \chi^2(n) \qquad (5.2.5c)$$

under the alternative hypothesis.

Using the results shown in Table 5.2.2, we can compute the LR test statistic.

$$\begin{aligned} LR &= 2[L(\theta_a) - L(\theta_0)] \\ &= 4399.10 - 1754.69 \\ &= 2644.41 > 12.20 \end{aligned}$$

The critical value from the chi-square distribution with degree of freedom of 26 at the 1 % significance level is 12.20 so the null hypothesis is rejected. That is, there is little statistical evidence to say that the constant coefficient model works better than the time-varying coefficient model.

5.3 Estimation of the Alternative Models of the Stock Returns

Let's consider alternative stock return generating structures. There have been proposed a lot of constant coefficient models. Out of them, OLS models, GARCH models and ARIMA models have been widely employed in many literatures. So the next thing to do is to estimate two models and compare their performance with that of the model in this thesis.

5.3.1. OLS Model of Stock Returns

Suppose that

$$y_t = x_t\beta + \epsilon_t$$

and $\epsilon_t \sim N(0, \sigma^2)$. Then the OLS estimator of β is a best linear unbiased estimated (BLUE).

Table 5.3.1. Results of Estimation of the OLS Model

Variable	Estimate	Standard Error	t-statistic
Constant	0.00450	0.00234	1.92649
TERM	1.35845	1.15867	1.17241
USPPI	0.14397	0.10839	1.32826
USIP	0.49024	0.06310	7.76904

The estimation results are summarized in the Table 5.3.1. None of the estimates of the coefficients on the macroeconomic variables except the growth rates of industrial production is statistically significant at the 5 % level. Positive correlation between the stock returns and the term spreads or the industrial production is compatible with results in the previous studies but that between the inflation rates and the stock returns is not. This simple model explains only 5.9 % of the variation in the stock returns. The actual stock returns and the estimates by the OLS model are plotted in Figure (5.7). As can be seen in that figure, the OLS estimates rarely follow the movement of the actual stock returns in magnitude and in direction. Thus its explanatory power is so low.

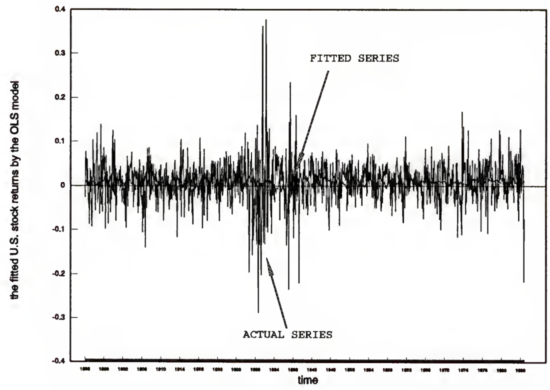

Figure 5.7 : The U.S. monthly stock returns and their fitted series by the OLS model using the U.S. monthly stock return indexes of the CRSP

5.3.2. GARCH(p,q) Models of Stock Returns

In generalized autoregressive conditional heteroskedasticity model (GARCH), first proposed by Engle (1982) and Bollerslev (1986), the current conditional variance is determined by squares of lagged estimated errors and lagged conditional variances. So far, there have been several versions of the GARCH models with different assumptions on the functional forms of the conditional variances. Since the GARCH model can produce good explanatory power over the series whose variances vary over time, there have been attempted a lot of applications to financial data such as stock market data and foreign exchange rate data.

The GARCH model to be estimated in this section is as follows:

$$y_t = C + \Gamma_1 TERM_t + \Gamma_2 USPPI_t + \Gamma_3 USIP_t + \theta\sqrt{h_t} + \epsilon_t, \quad \epsilon_t \sim N(0,\ h_t)$$

where

$$h_t = \alpha_0 + \sum_{i=1}^{3} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{3} \beta_j h_{t-j}$$

Table 5.3.2 shows that stock returns are insignificantly negatively related to the term spreads and positively to the inflation rates. Both results seem to contradict the results reported in the previous study in this area. This model also yields a statistically significantly positive correlation between the stock returns and the growth rates of industrial production. But current conditional variance appears to have no explanatory power over current stock return since the

estimate of θ is equal to zero. The estimates of the coefficients of the conditional variance equation imply that only one-month-lagged squared error has some significant effect on the current conditional variance. This model also yields a low R^2 of 0.0329.

Table 5.3.2 Results of Estimation of the GARCH(3,3) Model

Parameter	Estimate	Standard Error	t-statistic
C	.00636	.03854	.16489
TERM	-.78556	1.04670	-.75051
USPPI	.10547	.10252	1.02876
USIP	.47949	.05419	8.84806
θ	.00000	.77963	.00000
α_0	.00172	.01099	.15610
α_1	.00860	.00413	2.08373
α_2	.02441	.06674	.36573
α_3	.00375	.17334	.02163
β_1	.196653	7.51483	.02617
β_2	.000000	2.53755	.00000
β_3	.072707	.53582	.13569

The actual stock returns and the estimates by the GARCH(3,3) are plotted in Figure (5.8). We can find that like the OLS model, this model also does not produce good estimates following the actual series closely.

5.3.3. ARIMA(p,d,q) Models of Stock Returns

The general form of an autoregressive-integrated-moving average process of orders, p, d and q is defined as below:

$$\phi(L)\nabla^d y_t = \theta_0 + \theta(L)\varepsilon_t \qquad (5.3.3a)$$

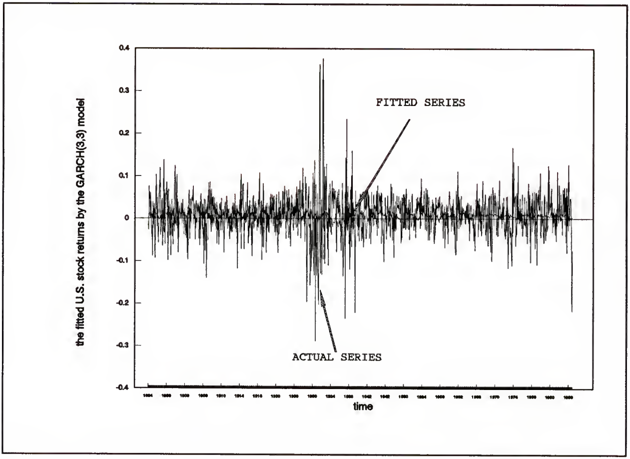

Figure 5.8 : The U.S. monthly stock returns and their fitted series by the GARCH(3,3) model using the U.S. monthly stock return indexes of the CRSP

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \tag{5.3.3b}$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \tag{5.3.3c}$$

where $\nabla^d y_t$ is the dth difference of y_t and L is a lag operator.

As the first step to estimation of the model, we have to choose appropriate values of p, d and q. In order to check whether $\{y_t\}$ is non-stationary or stationary, we need to do the so-called unit root test of $\{y_t\}$ because if $\{y_t\}$ is non-stationary, then regression will be spurious so that the differencing of $\{y_t\}$ should be made until it becomes stationary. The unit root test proposed by Dickey and Fuller (1979) is to regress ∇y_t on constant, time (t) and y_{t-1} and then to test the null hypothesis that the coefficient on y_{t-1} is equal to zero. That is,

$$\nabla y_t = \alpha_0 + \alpha_1 t + \alpha_2 y_{t-1} + e_t \tag{5.3.3d}$$

where $\nabla y_t = y_t - y_{t-1}$ and e_t is disturbance term. Then the testing hypotheses are given by

$H_0 : \alpha_2 = 0$

$H_a : \alpha_2 < 0$

Table 5.3.3a proves that the null hypothesis is rejected at 1 % of significance level since the test statistic is lower than the critical value of -29.5. It follows that the monthly stock returns are stationary in mean so that there is no need for differencing the stock return series. In other words, the appropriate value of d is equal to zero.

Next thing to do before estimating Equation (5.3.3a) is to

determine the suitable values of p and q. There are three

Table 5.3.3a Results of the Dickey- Fuller Test

Parameter	Estimate	Standard Error	t-Statistic
α_0	0.005600	0.003043	1.83897
α_1	0.000003	0.000004	0.77841
α_2	-0.908638	0.028986	-31.34750

most widely used criteria for selection of the most appropriate values of p and q. The earliest selection criterion is the Akaike Information Criterion (AIC), which is defined as

$$AIC(p,q) = \ln(S^2) + 2(p + q)/T$$

where S^2 is the estimate of the error variance σ^2 of the ARMA(p,q) model fitted to $z_t = \nabla^d y_t$. In general, however, the AIC tends to overparameterize the model. A second criterion, proposed by Russanen (1978) and Schwarz (1978), is given by

$$BIC(p,q) = \ln(S^2) + (p + q)T^{-1}\ln(T).$$

A third criterion, proposed by Hannan(1980), is defined as

$$PIC(p,q) = \ln(S^2) + (p + q)cT^{-1}\ln(\ln(T)).$$

Optimal values of p and q will be determined such that each of the criteria is minimized at these values. Hannan shows that, unlike the AIC, the BIC and the PIC are strongly consistent in that they determine the true model asymptotically. So in many cases, either the BIC or the PIC has been used in preference to the AIC.

Appendix 1 shows that both the BIC and the PIC are

minimized when p = 2 and q = 2 whereas the AIC is minimized when p = 8 and q = 2. We can find that these results are supportive to Hannan's argument. Combining the result of the unit root test with either the BIC or the PIC suggests that the ARIMA(2,0,2) model of the stock returns is most appropriate. That is, the regression equation is defined as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

and its estimation results are summarized in Table 5.3.3b.

Table 5.3.3b Results of Estimation of the ARIMA(2,0,2) Model

Parameter	Estimate	Standard Error	t-statistic
ϕ_1	-.382716	.088086	-4.34480
ϕ_2	-.655847	.078728	-8.33052
θ_1	-.508690	.077723	-6.54492
θ_2	-.760870	.065652	-11.5894

Table 5.3.3b shows that the current stock return is negatively correlated with two lagged stock returns and the shocks to the stock returns which appears over past two months. This model also yields very noisy estimates of the stock returns so the R^2 is equal to only 0.0349. In Figure (5.9), we can find that the estimates by the ARIMA(2,0,2) model move more volatilely than the actual stock returns do.

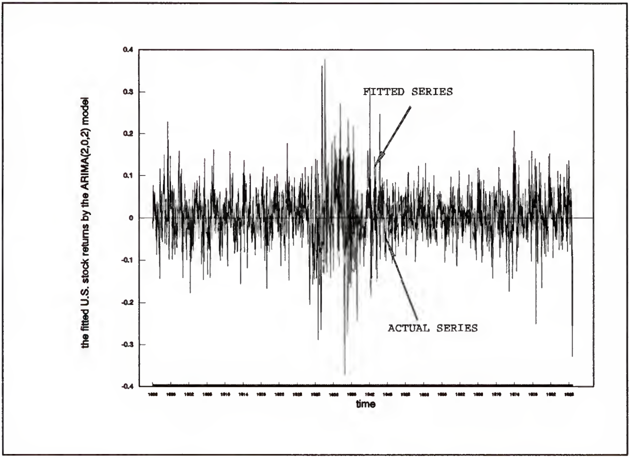

Figure 5.9 : The U.S. monthly stock returns and their fitted series by the ARIMA(2,0,2) using the U.S. monthly stock return indexes of the CRSP

5.3.4. Comparison between the Performances of the Varying Parameter Model and the Alternative Models

In this section, I will compare the overall performance of the time-varying parameter model of the stock returns with that of the alternative models of the stock returns.

As shown in Table (5.3.4), our model explains almost 93 % of the variation in the stock returns when current stock returns is in the known information and nearly 12.7 % of the variation in the stock returns when only the information about stock returns up to previous month is available. But none of the alternative models can explain more than 6 % of the stock return variation.

The value of the estimated mean squared error (MSE) is also commonly used as a measure of the model performance. The estimated mean squared errors calculated from our model is almost half of those obtained from the alternative models. Thus these two measures suggest that our model works better than the alternative models which assume time-invariant parameters.

Table 5.3.4. Comparison between the Alternative Models

Models	Estimated Mean Squared Errors	R^2
Kalman Filter Model		
Updating Equations	0.001393	0.92785
Prediction Equations	0.001374	0.12698
OLS Model	0.002587	0.05964
GARCH(3,3) Model	0.002598	0.05672
ARIMA(2,0,2) Model	0.002719	0.03290

5.4 Implications of the Kalman Filter Model on Dynamic Market Efficiency

In an efficient stock market, the asset prices of a specific firm fully and instantaneously reflect all available relevant information concerning the firm's business, and so do stock returns. It follows that an investor can make the more accurate expectation of a current stock return when he can have an access to some information concerning the firm's future business than when he relies just on the past history of the firm's business. However, once a new information on the firm's business is released, that will be fully and instantaneously incorporated into the firm's stock returns if stock markets are informationally efficient. So an expectation of current stock return based on the information available up to now should not be different from that based on the cumulative set of the information which already has been released.

In other words, a stock market is said to be dynamically efficient if $V(e_{t|t-1}) > V(e_t) = V(e_{t|T})$ for almost all $t \leq T$ where $e_{t|t-1} = y_t - x_t b_{t|t-1}$, $e_t = y_t - x_t b_t$ and $e_{t|T} = y_t - x_t b_{t|T}$.

The inequality between first two variances is an obvious result obtained from the Kalman filter.

$$e_t = y_t - x_t[b_{t|t-1} + P_{t|t-1}x_t'f_t^{-1}(y_t - x_t b_{t|t-1})]$$

$$= [1 - x_t P_{t|t-1}x_t'f_t^{-1}](y_t - x_t b_{t|t-1})$$

Note that $f_t = x_t P_{t|t-1}x_t' + h_t^c$, as defined in the Eq (4.2.3c),

is the variance of prediction error, s_t, so both terms are nonnegative definite. It follows that

$$0 < (1 - x_t P_{t|t-1} x_t' f_t^{-1}) < 1$$

Thus the variance of the term, e_t, is smaller than that of the other term, $e_{t|t-1}$.

Testing the dynamic efficiency of stock markets involves the joint test of the following three hypotheses:

Test 1:

$H_0 : V(e_t) = V(e_{t|t-1})$

$H_1 : V(e_t) < V(e_{t|t-1})$

Test 2:

$H_0 : V(e_{t|T}) = V(e_{t|t-1})$

$H_1 : V(e_{t|T}) < V(e_{t|t-1})$

Test 3:

$H_0 : V(e_t) = V(e_{t|T})$

$H_1 : V(e_t) > V(e_{t|T})$

For simplicity, let us assume that estimated errors, e_t, $e_{t|t-1}$ and $e_{t|T}$, are independently distributed with mean zero and that there exists f such that f_t converges to f as t goes to infinity. I perform the conventional F-test on those hypotheses. The test statistics of the tests are defined as $V(e_t)/V(e_{t|t-1})$, $V(e_{t|T})/V(e_{t|t-1})$ and $V(e_t)/V(e_{t|T})$, respectively. In the first two tests, the null is rejected if the test statistic is less than the critical value whereas in the third test the null is rejected if the opposite happens. Table (5.4) shows that in the first two tests the null hypotheses

are rejected but in the third test the null hypothesis is not rejected at 5 percents. Therefore we can conclude that there is found little evidence that stock markets are informationally inefficient.

Table 5.4 Results of Market Efficiency Tests

test	test statistic (F)	critical value	reject or not
Test 1	0.1427	0.9341	reject H_0
Test 2	0.2396	0.9341	reject H_0
Test 3	0.5955	1.0731	not reject H_0

CHAPTER 6
IMPLICATIONS ON THE CHARACTERISTICS OF THE STOCK RETURNS

6.1 Volatility of Stock Returns

First, our estimation results are very supportive for the fact that the volatility of stock returns varies over time, in particular, it tends to be generally very high during the war, economic recession, oil shocks, and banking panic. In this paper, volatilities of stock returns are measured by the squares of prediction errors. Figure 6.1 clearly shows that the volatilities of stock returns vary over time. For example, the high volatility of stock returns around the 1910s is because of the World War I, that around the 1930s is caused by the Great Depression, and that since the 1970s is ascribed to the oil shocks and the banking panic. Especially, the high volatilities which occurred around the Depression and the 1970s were very big in magnitude and lasted relatively for long time.

Second, our results also support the result, reported by Nelson (1989), French, Schwert, and Stambaugh (1987), or Poterba and Summers (1986), that shocks to volatility of stock returns are persistent with declining AR coefficients. As a test of persistence of shocks to volatility, I regress current

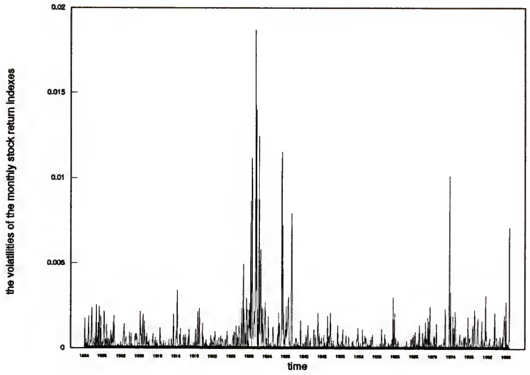

Figure 6.1 : The volatilities of the U.S. monthly stock returns measured by the squares of the estimated residuals of the prediction equations

squared prediction errors on the lagged squared prediction errors. That is,

$$\hat{\varepsilon}_t^2 = a_0 + a_1\hat{\varepsilon}_{t-1}^2 + a_2\hat{\varepsilon}_{t-2}^2 + a_3\hat{\varepsilon}_{t-3}^2 + a_4\hat{\varepsilon}_{t-4}^2 + a_5\hat{\varepsilon}_{t-5}^2 + a_6\hat{\varepsilon}_{t-6}^2 + a_{12}\hat{\varepsilon}_{t-12}^2 + a_{24}\hat{\varepsilon}_{t-24}^2 \tag{6.1.1a}$$

The regression results of the Eq (6.1.1a) are summarized in Table (6.1.1a). It is shown that the estimates of the AR

Table 6.1.1a Regression of the Volatility Measure

Variable	Estimated Coefficient	Standard Error	t-statistic
a_0	.000179	.000044	4.03852
a_1	.210394	.029938	7.02771
a_2	.094817	.031219	3.03715
a_3	.110118	.032430	3.39556
a_4	.022449	.031232	.71879
a_5	-.021027	.031091	-.67630
a_6	.011853	.030259	.39170
a_{12}	.159586	.031322	5.09507
a_{24}	.075936	.029374	2.58512

coefficients are declining with some seasonality because they smoothly decrease so the 4-month or further lagged volatilities become statistically negligible, but the coefficient estimate of the twelve month lagged volatility again becomes statistically significant. I find that it takes almost 0.44 months for a shock to volatility to become half as strong as the original one when a formula, $\rho^h = 1/2$, applies to the largest AR root in Table (6.1.1a). This appears to imply that the shocks to the volatility of stock returns fade out very quickly but they have some seasonality. Note that

Nelson (1989) shows that the half-life h of a shock associated with the largest root is about 7.3 years he obtains using daily data instead of monthly data. But this long persistence might stem from the misspecification of his model.

Third, low R^2 equal to 0.177 is obtained from the regression of the Eq (6.1.1a). This means that only a little portion of variation in the current volatility can be explained by the lagged ones.

6.2 Stock Returns and Economic Activity

First, our results in this work are supportive of the previous results, such as French, Schwert and Stambaugh (1987), Campbell and Shiller (1988), Fama and French (1989a,b), Turner, Startz and Nelson (1989), and Ball and Kothari (1989), that expected stock returns are time-varying. Two different types of expected stock returns are obtained from different sizes of information set and are plotted in Figure (5.5) and Figure (5.6), respectively. Figure (5.5) shows current expected stock returns based on the currently available information about stock returns while Figure (5.6) plots them based on one-month-lagged information. It is not difficult to find from these three figures that both series are varying over time.

Second, stock returns are, to some extent, predictable. Our model produces relatively high predictive power over stock returns. As shown in Table (5.3.4), nearly thirteen percent of

variation in monthly stock return index can be explained by the model employed in this work. Fama and French (1988a) argue that predictable variation of stock returns stems from slowly mean-reverting component of stock prices and then using individual firms' stock returns show that nearly 40 percent of variation of stock returns from portfolio of small firms can be predicted. But it may be worthwhile to note that their results are obtained from portfolio of much longer time horizon of 3 to 5 years than a month we assume in this work.

Third, as already shown in Section (5.2.2), relations between stock returns and macroeconomic variables such as interest rates, GNP growth rates, growth rates of domestic investment, and inflation rates are no longer stable. It can be easily found that the relationships such as a strong positive effect of changes in the industrial production, a positive effect of the term spreads and a negative effect of inflation rates on the stock returns are valid only for some sub-sample periods and are changing constantly over time.

Figure (5.1), Figure (5.2), Figure (5.3), and Figure (5.4) show how the estimates of the coefficients of the measurement equation vary over time. In Section (5.2.3), we find that the null hypothesis that parameters of our model do not vary over time is rejected at the 1 percent significance level.

6.3 Stock Returns and Volatility

In this section, I will analyze the relationship between the stock returns and their volatilities. First, I find that there exists a negative correlation between current stock return and next month volatility. To test it, I run a regression of current volatility of stock return on one-month-lagged stock return, that is,

$$VOLAT_t = \tau_0 + \tau_1 RET_{t-1} + e_t \qquad (6.3a)$$

where $VOLAT_t$, RET_{t-1} and e_t stand for the stock return volatility at time t, the stock return at time t-1 and disturbance term, respectively. The estimation results are given in Table (6.3a).

Table 6.3a Results of Estimation of the Eq (6.3a)

Variable	Estimated Coefficient	Standard Error	t-statistic
τ_0	0.00054	0.00129	0.42301
τ_1	-0.00249	0.00073	-3.41563

This supports the previous results, done by Nelson (1989d), Schwert (1989a) and Turner, Startz and Nelson (1989), that declines in the stock market tends to be associated with subsequent increases in volatility. Figure 6.2 shows the relationships between monthly stock return indexes and their volatilities.

Next, according to the mean variance theorem in finance, high risk associated with portfolios of risky financial assets

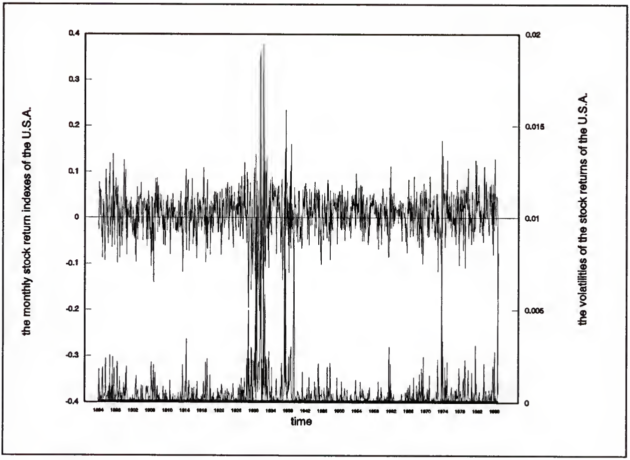

Figure 6.2 : The U.S. monthly stock returns and their estimated volatilities

should be compensated by high expected return from the portfolios. We can employ the volatility of the stock return index as a measure of the overall risk associated with investment in stock markets. Then we can derive a positive correlation between the expected return and its volatility. Since the current stock return is positively related to the expected return, the current stock return should be positively related to the volatility. Running a regression of the Eq (6.3b), I reassure this fact: refer to Table (6.3b).

$$RET_t = \Phi_0 + \Phi_1 VOLAT_t + e_t \tag{6.3b}$$

Table 6.3b Results of Estimation of the Eq (6.3b)

Variable	Estimated Coefficient	Standard Error	t-statistic
C	.00721	.00169	4.27596
VOLATILT	3.18468	1.21154	2.62861

CHAPTER 7
CONCLUDING REMARKS

The main purpose of this work is to construct a time series model which can explain the movement of stock returns well. One option to establish the purpose is to construct the stock return generating process, which depends on some macroeconomic variables such as the short-term interest rates, the inflation rates and the industrial production, and to allow the coefficients of linear regression model to vary with some degree of regularity over time by incorporating the Kalman filter into the linear model. Before constructing the Kalman filter model, I thought that it is first necessary to clarify the distribution of stock returns since the previous work points out that stock returns tend to have fatter tails and bigger kurtosis than the normal. Out of several density estimation methods, the naive estimator and kernel estimator are obtained with different estimation procedures from the previous ones. I find that the difference between the normal and the so-called generalized error distribution is indistinguishable in capturing the true distribution and that the GED requires higher computational cost than the normal distribution. Thus the normal distribution is employed in

this work as the distribution of monthly stock returns.

In Chapter 4, I construct a more general form of the Kalman filter model allowing the disturbance terms in both the measurement equation and the transition equation to follow an ARCH(1) process. It is shown that the ARCH(1) assumption yields different forms of the prediction equation and the updating equation.

In Chapter 5, a simpler form than the described one in the previous chapter is estimated by the Davidon-Fletcher-Powell iteration method since the general form of the time-varying parameter model needs many unknown parameters to be estimated simultaneously so that it causes the tremendous computational costs. Based on the estimation results, I perform two tests, the test of the time-varying model against the time-invariant model and the test of stock market efficiency. At the first test, the null hypothesis that the true model of the stock returns is time-invariant is rejected at 1 % of significance level. At the second test, the null hypothesis that the stock markets are informationally efficient is not rejected at 5 %.

And I compare the performance of the model employed in this thesis with that of the alternative models such as the OLS model, the GARCH(3, 3) model and the ARIMA(2, 0, 2) model and find that our model has better explanatory power for the variation of the monthly CRSP stock returns than the alternative models do. Finally, based on the results of the estimation of the Kalman filter model, the characteristics of

the U.S. stock returns, which are found in the previous work, are reinterpreted. I find that many of them reach different implications.

APPENDIX

Table A.1 Criteria for Choosing the Appropriate Values of p and q

NAR	NMA	0	1	2	3	4	5	6
1	AIC	-5.8843	-5.8819	-5.8803	-5.8880	-5.8863	-5.8979	-5.8959
	BIC	-5.8800	-5.8734	-5.8674	-5.8709	-5.8649	-5.8722	-5.8660
	PIC	-5.8827	-5.8787	-5.8754	-5.8816	-5.8783	-5.8882	-5.8846
2	AIC	-5.8822	-5.8805	-5.9009	-5.8984	-5.8961	-5.8998	-5.8981
	BIC	-5.8737	-5.8677	-5.8838	-5.8770	-5.8705	-5.8698	-5.8639
	PIC	-5.8790	-5.8757	-5.8944	-5.8904	-5.8865	-5.8885	-5.8852
3	AIC	-5.8860	-5.8845	-5.8984	-5.8959	-5.8942	-5.8982	-5.8965
	BIC	-5.8732	-5.8674	-5.8770	-5.8702	-5.8643	-5.8639	-5.8580
	PIC	-5.8812	-5.8780	-5.8904	-5.8862	-5.8829	-5.8853	-5.8820
4	AIC	-5.8873	-5.8856	-5.8961	-5.8944	-5.8927	-4.6907	-4.6890
	BIC	-5.8702	-5.8642	-5.8704	-5.8645	-5.8585	-4.6522	-4.6462
	PIC	-5.8809	-5.8776	-5.8864	-5.8831	-5.8798	-4.6762	-4.6729
5	AIC	-5.8964	-5.8969	-5.8956	-5.8978	-4.9850	-5.0890	-5.0873
	BIC	-5.8750	-5.8712	-5.8657	-5.8636	-4.9465	-5.0462	-5.0402
	PIC	-5.8883	-5.8872	-5.8844	-5.8849	-4.9705	-5.0728	-5.0695
6	AIC	-5.8964	-5.8946	-5.8929	-5.8912	-5.0539	-5.0522	-5.0505
	BIC	-5.8708	-5.8647	-5.8587	-5.8527	-5.0111	-5.0052	-4.9992
	PIC	-5.8868	-5.8833	-5.8800	-5.8767	-5.0378	-5.0345	-5.0312
7	AIC	-5.8965	-4.6121	-4.6104	-5.9033	-5.9022	-5.8997	-5.9013
	BIC	-5.8666	-4.5778	-4.5719	-5.8605	-5.8551	-5.8483	-5.8456
	PIC	-5.8853	-4.5992	-4.5959	-5.8872	-5.8844	-5.8803	-5.8803
8	AIC	-5.9003	-5.8994	-5.9040	-5.9023	-5.9006	-5.8989	-5.8972
	BIC	-5.8661	-5.8609	-5.8612	-5.8552	-5.8492	-5.8433	-5.8373
	PIC	-5.8874	-5.8849	-5.8878	-5.8845	-5.8812	-5.8779	-5.8746
9	AIC	-5.8993	-5.8969	-5.8952	-5.8935	-5.1419	-5.1402	-5.1385
	BIC	-5.8607	-5.8541	-5.8482	-5.8422	-5.0863	-5.0803	-5.0743
	PIC	-5.8847	-5.8808	-5.8775	-5.8742	-5.1209	-5.1176	-5.1143

(Continued)

NAR	NMA	7	8	9	10	11	12
1	AIC	-5.8915	-5.8911	-5.8981	-5.8964	-5.8932	-5.8915
	BIC	-5.8573	-5.8526	-5.8553	-5.8494	-5.8418	-5.8358
	PIC	-5.8786	-5.8766	-5.8820	-5.8787	-5.8738	-5.8705
2	AIC	-5.9058	-4.9586	-5.8956			
	BIC	-5.8673	-4.9159	-5.8485			
	PIC	-5.8913	-4.9425	-5.8779			
3	AIC	-5.0550	-5.9012	-5.9036	-5.9019	-5.9003	-5.8986
	BIC	-5.0122	-5.8541	-5.8523	-5.8463	-5.8403	-5.8344
	PIC	-5.0389	-5.8834	-5.8843	-5.8810	-5.8777	-5.8744
4	AIC	-5.9062	-5.9040	-3.5525	-5.8879	-5.8863	-5.8846
	BIC	-5.8591	-5.8526	-3.4969	-5.8280	-5.8221	-5.8161
	PIC	-5.8884	-5.8846	-3.5315	-5.8654	-5.8621	-5.8588
5	AIC	-5.8924	-5.8907	-5.8890	-2.6152	-2.6135	-2.6118
	BIC	-5.8411	-5.8351	-5.8291	-2.5510	-2.5451	-2.5391
	PIC	-5.8731	-5.8698	-5.8665	-2.5910	-2.5877	-2.5844
6	AIC	-5.8851	-5.8896	-4.0352	-4.0335	-4.0319	-4.0302
	BIC	-5.8295	-5.8297	-3.9710	-3.9651	-3.9591	-3.9531
	PIC	-5.8642	-5.8670	-4.0110	-4.0077	-4.0044	-4.0011

AIC : the Akaike Information Criterion
BIC : the Schwarz Information Criterion
PIC : the Hannan Information Criterion
NMA : the Number of Moving Average Terms
NAR : the Number of Autoregressive Terms

BIBLIOGRAPHY

Ansley, C. F. & Kohn, R. (1982) "A Geometrical Derivation of the Fixed Interval Smoothing Algorithm," Biometrika 69, 486-487.

Ansley, C. F. & Kohn, R. (1985) "Estimation, Filtering, and Smoothing in State Space Models with Incompletely Specified Initial Conditions," The Annals of Statistics, Vol. 13, No. 4, 1286-1316.

Aoki, M. (1987) "A State Space Time Series Modeling Method without Individual Detrending," Journal of Time Series Analysis, Vol. 12, No. 1.

Ball, R. and Kothari, S. P. (1989) "Nonstationary Expected Returns Implications for Tests of Market Efficiency and Serial Correlation in Returns," Journal of Financial Economics 25, 51-74.

Black, Fischer (1976) "Studies of Stock Price Volatility Changes," Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association, 177-181.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroskedasticity," Journal of Econometrics, 31, 307-327.

Bollerslev, T. (1987) "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," Review of Economics and Statistics, 69, 542-547.

Bollerslev, T. (1988) "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process," Journal of Time Series Analysis, 9, 121-131.

Bollerslev, T., Chou, R. Y., Jayaraman, N. and Kroner, K. F. (1990) "ARCH Modeling in Finance A Selective Review of the Theory and Empirical Evidence, with Suggestions for Further Research," Working Paper, Northwestern University.

Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1987) "A Capital Asset Pricing Model with Time Varying Covariances," Journal of Political Economy, 96, 1, 116-131.

Boneva, L. I., Kendall, D. G. and Stefanov, I. (1971) "Spline Transformations: Three New Diagnostic Aids for the Statistical Data-Analysis (with Discussion)," Journal of the Royal Statistical Society B, 33, 1-70.

Breiman, L., Meisel, W. and Purcell, E. (1977) "Variable Kernel Estimates of Multivariate Densities," Technometrics, 19, 135-144.

Campbell, J. Y. (1990) "Measuring the Persistence of Expected Returns," American Economic Review, Vol. 80, No. 2, 43-47.

Campbell, J. Y and Hentschel, L. (1990) "No news is good news An Asymmetric model of changing Volatility in stock returns," Working Paper, Princeton University and National Bureau of Economic Research.

Campbell, J. Y. and Shiller, R. J. (1988) "Stock Prices, Earnings, and Expected Dividends," Journal of Finance, Vol. 43, 661-676.

Cencov, N. N. (1962) "Evaluation of an Unknown Distribution Density from Observations," Soviet Mathematics, 3, 1559-1562.

Chamberlain, G. (1983) "Funds, Factors, and Diversification in Arbitrage Pricing Models," Econometrica, Vol. 51, No. 5.

Chamberlain, G. and Rothschild, M. (1983) "Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets," Econometrica, Vol. 51, No. 5, 1281-1304.

Chen, N. (1984) "Some Empirical Tests of the Theory of Arbitrage Pricing," Journal of Finance Vol. 38, 1393-1414.

Chow, G.C. (1984) "Random and Changing Coefficient Models," Handbook of Econometrics, Ch. 21, North-Holland, New York.

Connor, G. (1980) "Asset Prices in a Well-Diversified Economy," Technical Report No. 47, Yale School of Organization and Management.

Copas, J. B. and Fryer, M. J. (1980) "Density Estimation and Suicide Risks in Psychiatric Treatment," Journal of the Royal Statistical Society A, 143, 167-176.

de Jong, P. (1988) "The Likelihood for a State Space Model," Biometrika 75, 165-169.

de Jong, P. (1989) "The Diffuse Kalman Filter," working paper, University of British Columbia.

de Jong, P. (1991) "Stable Algorithms for the State Space Model," Journal of Time Series Analysis, Vol. 12, No. 2, 143-157.

Dhrymes, P. J., Friend, I. and Gultekin, N. B. (1984) "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory," Journal of Finance, Vol. 39, 323-346.

Dhrymes, P. J., Friend, I., Gultekin, M. N. and Gultekin, N. B. (1985) "New Tests of the APT and Their Implications," Journal of Finance, Vol. 40, 659-675.

Engle, R. F. (1982) "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," Econometrica, Vol. 50, 987-1007.

Engle, R.F. and Bollerslev, T. (1986a) "Modeling the Persistence of Conditional Variance," Econometric Reviews, 5, 1, 1-50.

Engle, R. F. and Bollerslev, T. (1986b) "Modeling the Persistence of Conditional Variances A Reply," Econometric Reviews, 5, 1, 81-87.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987) "Estimating Time Varying Risk Premia in the Term Structure The ARCH-M Model," Econometrica, 55, 2, 391-408.

Fama, E. F. (1990) "Stock Returns, Expected Returns, and Real Activity," Journal of Finance, Vol. 45, No. 4., 1089-1108.

Fama, E. F. and French, K. R. (1988a) "Permanent and Transitory Components of Stock Prices," Journal of Political Economy 96, 246-273.

Fama, E. F. and French, K. R. (1988b) "Dividend Yields and Expected Stock Returns," Journal of Financial Economics 17, 3-25.

Fix, E. and Hodges, J. L. (1951) "Discriminatory Analysis, Nonparametric Estimation: Consistency Properties," Report No. 4, Project No. 21-49-004, USAF School of Aviation Medichine, Randolph Field, Texas.

French, K. R. and Roll, R. (1986) "Stock Return Variances: The Arrival of Information and the Reaction of Traders," Journal of Financial Economics 17, 5-26.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987) "Expected Stock Returns and Volatility," Journal of Financial Economics, 19, 1, 3-29.

Geweke, J. (1986) "Modeling the Persistence of Conditional Variances A Comment," Econometric Reviews, 5, 1, 57-61.

Good, I. J. and Gaskins, R. A. (1971) "Nonparametric Roughness Penalties for Probability Densities," Biometrika, 58, 255-277.

Harvey, A. C. (1989) "Forecasting. Structural Models and the Kalman Filter," Cambridge University Press, New York.

Harvey, A. C. and Snyder, R. D. (1990) "Structural Time Series Models in Inventory Control," International Journal of Forecasting 6, 187-198.

Harvey, A. C. and Ruiz, E. (1990) "Unobserved Component Time Series Models with ARCH Disturbances," Working Paper, London School of Economics.

Hong, C. H. (1988) "The Integrated Generalized Autoregressive Conditional Heteroskedastic Model: The Process, Estimation and Monte Carlo Experiment," mimio, University of California at San Diego.

Kaul, G. and Seyhun, H. N. (1990) "Relative Price Variability, Real Shocks, and the Stock Market," Journal of Finance, Vol. 45, No. 2, 479-96.

Kearns, P. and Pagan, A. R. (1990) "Australian Stock Market Volatility 1875-1987," Working Paper, University of Rochester.

Lebaron, B. (1990) "Some Relations Between Volatility and Some Correlations in Stock Market Returns," Working Paper, University of Wisconsin-Madison.

Litner, J. (1965) "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," Review of Economics and Statistics, February, 13-37.

Loftsgaarden, D. O. and Quesenberry, C. P. (1965) "A Non-Parametric Estimate of a Multivariate Density Function," The Annals of Mathematical Statistics, 36, 1049-1051.

Mei, J. (1989) "Variable-Expected-Returns and the Present Value Model: A Panel Study," Working Paper, Princeton University.

Merville, L. J. and Pieptea, D. R. (1989) "Stock Price Volatility, Mean reverting diffusion, and noise," Journal of Financial Economics 24, 193-214.

Mossin, J. (1966) "Equilibrium in a Capital Asset Market," Econometrica, October, 768-783.

Nelson, D. B. (1988a) "The Time Series Behavior of Stock Market Volatility and Returns," Unpublished Doctoral Dissertation, Massachusetts Institute of Technology.

Nelson, D. B. (1988b) "Stationarity and Persistence in the GARCH(1,1) Model," Working Paper, #88-68, University of Chicago.

Nelson, D. B. (1988c) "ARCH Models as Diffusion Approximations," Working Paper, #88-69, University of Chicago.

Nelson, D. B. (1989) "Conditional Heteroskedasticity in Asset Returns A New Approach," Working Paper, University of Chicago.

Newey, W. K. (1985a) "Maximum Likelihood Specification Testing and Conditional Moment Tests," Econometrica, 53, 1047-1070.

Newey, W. K. (1985b) "Generalized Method of Moments Specification Testing," Journal of Econometrics, 29, 3, 229-256.

Pagan, A. R. and Hong, Y. S. (1990) "Nonparametric Estimation and the Risk Premium," in W. Barnett, J. Powell and G. Tauchen, eds, Semiparametric and Nonparametric Methods in Econometrics and Statistics, Cambridge University Press, Cambridge.

Pagan, A. R. and Sabau, H. C. L. (1987) "Consistency Tests for Heteroskedastic and Risk Models," Mimeo, University of Rochester Economics Department, October.

Poterba, J.M and Summers, L. H. (1986) "The Persistence of Volatility and Stock Market Fluctuations," American Economic Review, 76, 5, 1142-1151.

Poterba, James M., and Summers, L. H. (1987) "Mean Reversion in Stock Returns Evidence and Implications," Mimeo, National Bureau of Economic Research, Cambridge.

Reinganum, M. R. (1981) "Empirical Tests of Multi-factor Pricing Model," Journal of Finance, Vol. 36, 313-321.

Robinson, P. M.(1991) "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression," Journal of Econometrics 47, 67-84.

Roll, R. (1977) "A Critique of the Asset Pricing Theory's Tests," Journal of Financial Economics, March, 129-176.

Roll, R. (1981) "A Possible Explanation of the Small Firm Effect," Journal of Finance, Vol. 36, 879-888.

Roll, R. and Ross, S. A. (1980) "An Empirical Investigation of the Arbitrage Pricing Theory," Journal of Finance, Vol. 35, 1073-1103.

Rosenblatt, M. (1956) "Remarks on Some Nonparametric Estimates of a Density Function," Annals of Mathematical Statistics, 27, 832-837.

Ross, S. (1976) The Arbitrage Theory of Capital Asset Pricing," Journal of Economic Theory, 341-360.

Ross, S. (1980) "An Empirical Investigation of the Arbitrage Pricing Theory," Journal of Finance, Vol. 35, 1073-1103.

Sawyer, K. (1989) "Testing the Bounding Conditions of Arbitrage Pricing," Working Paper, University of Florida.

Schwert, G. W. (1989a) "Why Does Stock Market Volatility Change Over Time?," Journal of Finance, Vol. 44, 1115-1153.

Schwert, G. W. (1989b) "Business Cycles, Financial Crises and Stock Volatility," Carnegie-Rochester Conference Series on Public Policy 31, 83-125.

Schwert, G. W. (1989c) "Stock Market Volatility," Working Paper #89-24, University of Rochester.

Shanken, J. (1982) "The Arbitrage Pricing Theory Is it Testable?," Journal of Finance, Vol. 37, 1129-40.

Shanken, J. and Weinstein, M. I. (1990) "Macroeconomic Variables and Asset Pricing Further Results," Working Paper, University of Rochester.

Sharpe, W. F. (1964) "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," Journal of Finance, Vol. 19, 425-442

Turner, C.M., Startz, R., and Nelson, C.R. (1989) "A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market," Journal of Financial Economics 25, 3-22.

Wei, K. (1988) "An Asset-Pricing Theory Unifying the CAPM and APT," Journal of Finance, Vol. 43, No. 4, 881-892.

White, H. (1980) "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica 48, 817-38.

Whittle, P. (1958) "On the Smoothing of Probability Density Functions," Journal of the Royal Statistical Society B, 20 334-343.

BIOGRAPHICAL SKETCH

Young-Hoon Koo was born in Kyungnam, Korea, on February 5, 1959. In 1985, he received a B. A. degree in economics from the Pusan National University, Pusan, Korea. In 1988, after he graduated from the M. A. program in economics of the same university, he entered the doctoral program in economics at the University of Florida and anticipate the completion of his doctoral studies in August of 1992.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

G. S. Maddala, Chairman
Graduate Research Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David A. Denslow, Jr.
Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mark Rush
Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert D. Emerson
Professor of Food and Resource Economics

This dissertation was submitted to the Graduate Faculty of the Department of Economics in the College of Business Administration and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1992

Dean, Graduate School